

Konseptvalgutredning

Ny personidentifikator i Folkeregisteret

4. april 2013

Innhold

1	Sammendrag.....	5
2	Innledning.....	9
2.1	Om Folkeregisteret	9
2.2	Problemstilling	10
2.3	Mandat.....	10
2.4	Avgrensning og rammer for utredningen	11
2.5	Grensesnitt og føringer.....	11
2.6	Organisering	12
2.7	Utforming av KVUen.....	13
2.8	Referansedokumenter	13
3	Behovsanalyse	14
3.1	Dagens fødselsnummer	14
3.2	Andre personnumre i bruk.....	17
3.3	Andre land	19
3.4	Normative behov	19
3.5	Etterspørselsbaserte behov	21
3.6	Interessentgruppebaserte behov	25
3.7	Prosjektutløsende behov.....	31
3.8	Oppsummering av behov.....	32
4	Strategi	34
4.1	Innledning.....	34
4.2	Samfunns målet for tiltaket	34
4.2.1	Effektmål.....	35
4.2.2	Målkonflikter.....	36
5	Krav	37
5.1	Innledning.....	37
5.1.1	Absolutt krav:.....	37
5.1.2	Viktige krav.....	38
6	Mulighetsstudie	39
6.1	Gjenbruk av fødselsnumre fra døde personer.....	39
6.2	Mulighetsdimensjoner	39
6.2.1	Alfanumerisk.....	39
6.2.2	Andre egenskaper enn numre/bokstaver	40
6.2.3	Biometri.....	40
6.2.4	Informasjonsløs identifikator	40
6.2.5	Antall kontrollsifre	40
6.2.6	Sektorvis identifikator	40
6.2.7	Konvertering	40
6.2.8	Lengde.....	40
6.3	Vurdering av mulighetsdimensjoner	41
6.3.1	Alfanumerisk.....	41
6.3.2	Andre egenskaper enn numre/bokstaver	41
6.3.3	Biometri.....	41
6.3.4	Informasjonsløs identifikator	42
6.3.5	Antall kontrollsifre	42
6.3.6	Sektorvis identifikator	42
6.3.7	Konvertering	42

	6.3.8	Lengde.....	42
	6.4	Oppsummering	42
7		Alternativanalyse.....	44
	7.1	Beskrivelse av alternativene	44
	7.1.1	Nullalternativet – dagens fødselsnummer	44
	7.1.2	Alternativ 1 - Dagens fødselsnummer med endret kontrollsiffer	45
	7.1.3	Alternativ 2 – Alfanumerisk og informasjonsbærende.....	46
	7.1.4	Alternativ 3 – Informasjonsløst kundennummer	48
	7.1.5	Alternativ 4 – Alfanumerisk kundennummer med åtte siffer	49
	7.2	Robusthetsanalyse	49
	7.2.1	Kapasitet for dagens populasjon.....	50
	7.2.2	Robusthet for utvidelse av dagens populasjon	51
	7.3	Prosjektperiode	53
	7.4	Metode for den samfunnsøkonomiske analysen.....	54
	7.4.1	Datainnsamling	55
	7.4.2	Forutsetninger for den samfunnsøkonomiske analysen.....	55
	7.5	Nytte- og kostnadsestimater i analysen	56
	7.5.1	Merkostnader.....	56
	7.5.2	Kostnadselementer	56
	7.5.3	Inndeling i kostnadsgrupper	57
	7.5.4	Fordeling av kostnader mellom sektorer	58
	7.5.5	Nytte.....	59
	7.6	Usikkerhetsanalyse.....	59
	7.6.1	Resultater fra usikkerhetsanalysen	59
	7.6.2	Effekt av periodisering og diskontering	60
	7.7	Ikke prissatte effekter	61
	7.7.1	Ivareta et godt personvern	62
	7.7.2	Brukervennlig personidentifikator	63
	7.7.3	Sannsynligheten for feilregistrering	64
	7.7.4	Robusthet mot endringer av personopplysninger	66
	7.7.5	Sameksistens med dagens fødselsnummer og andre brukte nummerserier	68
	7.7.6	Oppsummering av ikke-prissatte virkninger for de ulike alternativene.....	69
	7.8	Vurdering av fordelingseffekter	69
	7.9	Vurdering av realopsjoner	70
	7.9.1	Realopsjoner for nullalternativet	70
	7.9.2	Realopsjoner mellom alternativene	70
	7.10	Sensitivitetsanalyse	70
	7.10.1	Diskontering.....	70
	7.10.2	Diskonteringsrente	70
	7.10.3	Restverdi	70
	7.10.4	Størrelse på uspesifisert.....	71
	7.11	Samlet samfunnsøkonomisk analyse.....	72
	7.12	Finansieringsplan.....	73
	7.13	Hovedkonklusjon.....	73
8		Føringer for forprosjektfasen.....	74
	8.1	Innledning.....	74
	8.2	Viktige elementer fra analysen	74
	8.2.1	Behovsanalyse.....	75
	8.2.2	Kravdokument.....	75
	8.2.3	Gjennomføringsstrategi	75
	8.3	Koordinert samhandling mellom virksomhetene	76
	8.4	Overgangsordninger	77
	8.5	Kontraktstrategi	77
	8.6	Suksessfaktorer og fallgruver	78
	8.6.1	Tidsløpet.....	78
	8.6.2	God eierstyring.....	79

8.6.3	Erfaringer fra tidligere prosjekter	79
8.7	Usikkerheter	79
8.8	Gevinstrealiseringsplan	80
8.9	Prosjektorganisering.....	80
Bibliografi		81
Vedlegg		82

1 Sammendrag

Bruken av Det sentrale folkeregister har økt i takt med samfunnsutviklingen. Registeret har blitt sentralt for nærmest all offentlig personrelatert saksbehandling og planlegging. Også for store deler av privat sektor er Folkeregisteret en nødvendig informasjonskilde for å utføre lovpålagte oppgaver.

Med utvikling av IKT og elektronisk kommunikasjon har et korrekt og oppdatert folkeregister bidratt til store effektivitetsgevinster for hele det norske samfunn. Folkeregisteret har blitt en nødvendig del av vår felles infrastruktur og brukes som "primærnøkkel" for samhandling mellom virksomheter.

Finansdepartementet har gitt Skattedirektoratet i oppdrag å utrede nærmere de samfunnsmessige, økonomiske og administrative konsekvenser knyttet til valg av ny personidentifikator. Prosjektet sees i sammenheng med strategi for nytt folkeregister og inngår i Skattedirektoratets *Program for modernisering av Folkeregisteret*.

Utredningens mandat angir at det utarbeides en konseptvalgutredning (KVU) i henhold til de krav og føringer som er stilt til KVU i Finansdepartementets rammeverk for kvalitetssikring.

Behovsanalyse

Følgende er de viktigste behovene satt opp i prioritert rekkefølge:

1. Kapasiteten til dagens fødselsnummer er beregnet å bli for liten fra og med 1. januar 2040 og må derfor innen dette tidspunkt være erstattet av en ny personidentifikator med tilstrekkelig kapasitet. Dette er tiltakets prosjekttløsende behov.
2. Med dagens regelverk og populasjon må en ny personidentifikator ha følgende kapasitet:
 - Informasjonsbærende: Minst 900 ulike verdier per dato i 2150.
 - Informasjonsløs: 37 millioner unike identifikatorverdier fram til 2150.
3. Stadig økende bruk av Folkeregisteret, samt fremvekst av elektroniske tjenester og elektronisk samhandling, setter store krav til kvaliteten på opplysninger i Folkeregisteret og i særdeleshet på personidentifikatoren.

Mandatet for KVUen legger til grunn dagens lover og regler for definisjon av populasjonen.

Usikkerhet i forutsetninger, prognoser og mulighet for at spesielle hendelser inntreffer, tilsier gode kapasitetsmarginer for å dekke eventuelt økt behov.

Det er registrert interessekonflikt knyttet til informasjonsløs kontra informasjonsbærende personidentifikator. Hovedargumentene for de to typer identifikatorer er:

Informasjonsløs personidentifikator:

- Økt stabilitet og robusthet for endringer i persondata
- Høy kapasitet
- Prinsipielt uheldig at identifikatoren inneholder informasjon

Informasjonsbærende personidentifikator:

- Brukervennlig og lett å huske
- Redusere antall feil
- Mer kostnadseffektiv
- Gir raskere tilgang til vitale personopplysninger som alder og kjønn
- Mindre endring fra dagens system

Ved innføring av ny personidentifikator er det behov for en hensiktsmessig overgangsordning som minimerer kostnader og ulemper for virksomheter og privatpersoner. Hvis overgangen til ny identifikator innebærer systemendringer av litt størrelse, bør overgangsperioden være minst 10 år.

Det er identifisert et behov for å se fødselsnummeret i sammenheng med D-nummer (se avsnitt 3.2 om D-nummerordningen). Det pågående *Program for modernisering av Folkeregisteret* vurderer mulighetene for å utvide dagens populasjon i Folkeregisteret til også å inkludere D-numre. Hvis beslutningsprosessen her resulterer i at populasjonen skal endres, vil det medføre økte krav til kapasitet for ny personidentifikator.

Ved utvikling av nye databaser anbefales benyttet databasegenererte primærnøkler, mens personidentifikatoren legges i et informasjonsfelt som er designet for oppslag, søk etc. og med robusthet for fremtidige endringer.

Samfunns mål og effektmål

Tiltaket ny personidentifikator i Folkeregisteret har følgende samfunns mål:

En unik personidentifikator egnet for praktisk anvendelse til minst år 2150.

Tiltaket ny personidentifikator i Folkeregisteret har følgende effektmål:

1. Tilstrekkelig kapasitet for dagens registerpopulasjon og robusthet for eventuell framtidig utvidelse av populasjonen.
2. Kostnadseffektiv realisering og drift.
3. Ha en utforming som oppleves brukervennlig.

Krav

Det stilles tre absolutte krav til ny personidentifikator, innenfor kapasitet og personvern:

1. Ha kapasitet for dagens registerpopulasjon minst fram til 2150.
2. Hver person skal ha en unik personidentifikator.
3. Oppfylle gjeldende regelverk for personvern og sikkerhet.

I tillegg er det inkludert fire viktige krav som er vurdert som sentrale for at vi skal få et velfungerende Folkeregister med ny personidentifikator. De er:

1. Lav gjennomføringsrisiko.
2. Høy brukervennlighet.
3. Gode kontrollrutiner.
4. God sameksistens med andre nummerserier.

Mulighetsrommet

Følgende mulighetsdimensjoner vurderes som realistiske og videreføres i alternativanalysen:

- Alfanymerisk eller numerisk
- Informasjonsløs eller informasjonsbærende identifikator

- Beregning av kontrollsifre
- Med og uten konvertering.
- Lengde

Ut i fra disse dimensjonene er det gjort en prosess med grovsiling av alternativer. Man startet med over 40 alternativer, og følgende fire hovedalternativer er gjenstand for analyse.

	Informasjonsbærende	Informasjonsløs
Numeriske tegn	0. Dagens fødselsnummer: DDMMÅÅIIKK. 1. Dagens løsning med endret kontrollberegning: DDMMÅÅIIKK.	3. Kundenummer: IIIIIIIKK. Har to varianter: 3a) Kun nye borgere 3b) Hele befolkningen skifter.
Alfanumeriske tegn	2. Informasjonsbærende og alfanumerisk: DDMMÅÅAAKK.	4. Informasjonsløs og alfanumerisk (8): AAAAAAKK. Hele befolkningen skifter.

Alternativanalyse

Alternativanalysen omfatter:

1. Robusthetsanalyse av alternativenes kapasitet.
2. Samfunnsøkonomisk analyse med prissatte- og ikke prissatte-konsekvenser.

1. Robusthetsanalysen

Alternativenes kapasitet analyseres i forhold til to ulike utviklingsløp for framtidig personidentifikator:

- a) Alternativet hvor dagens kriterier for tildeling av fødselsnummer beholdes.
- b) Alternativet hvor dagens kriterier for tildeling av fødselsnummer endres/utvides til å inkludere dagens D-nummer populasjon.

Robusthetsanalysen viser at dagens fødselsnummer (nullalternativet) ikke har tilstrekkelig kapasitet og kravet om unik identifisering kan opphøre fra og med 2040. Dette er et absolutt krav og nullalternativet utgår derfor som framtidig personidentifikator for nye innbyggere.

De fire alternativene (1-4) oppfyller kravene til kapasitet med god margin for alternativet med dagens kriterier for tildeling av fødselsnummer. For alternativet hvor D-nummer inngår i en framtidig personidentifikator, vil med stor sannsynlighet alle fire alternativene oppfylle kapasitetskravet.

2. Samfunnsøkonomisk analyse med prissatte- og ikke prissatte-konsekvenser

Det anbefalte konseptet er Alternativ 1. Dette har langt lavere kostnader enn samtlige andre alternativer og med håndterbar risiko. Alternativ 1 er bedre enn Alternativ 2 blant de

informasjonsbærende, dette som følge av langt høyere kostnader for Alternativ 2 og kun beskjedne forbedringer på ikke-prissatte effekter.

Alternativene med full konvertering av eksisterende fødselsnumre er svært kostbare og har svært høy risiko. I tillegg til risikoen uttrykt ved et høyt kostnadsmessig standardavvik, har flere virksomheter stilt spørsmålsteget ved om prosjektet er gjennomførbart. Dessuten er det varslet at det kan få konsekvenser for liv og helse.

Alternativ 3a, som ikke innebærer konvertering, er eneste gjenværende alternativ som er informasjonsløst. Noen virksomheter foretrekker et slikt alternativ, men kostnaden er estimert til anslagsvis 4 milliarder kroner høyere enn for Alternativ 1. De samlede ikke-prissatte virkningene er vurdert som marginalt dårligere enn for Alternativ 1.

Føringer for forprosjektfasen

Det er det nasjonale "fellesprosjektet" med å innføre ny personidentifikator som er gjenstand for omtale i dette kapitlet. I vurderingen blir Skattedirektoratets muligheter til å lede og koordinere et nasjonalt prosjekt avgjørende for suksess. Det kan oppstå konkrete problemstillinger hvor det er viktig at Skattedirektoratet som prosjekteier har fokus på fellesskapet framfor egne interesser.

I tillegg til det nasjonale fellesprosjektet vil det pågå flere virksomhetsinterne prosjekter i berørte virksomheter. Det er ikke Skattedirektoratets ansvar å styre disse prosjektene, og vi har i dette kapitlet ikke vurdert hvordan de enkelte prosjektene i et utall virksomheter bør organiseres og styres.

Det er en betydelig iboende institusjonell usikkerhet ved at vesentlige deler av aktiviteten skjer fragmentert i et utall virksomheter, private og offentlige. Dette er en situasjon man ikke kan oppløse og som er utenfor prosjekteiers kontroll. Dette gir økt risiko. Sentrale grep blir informasjon, motivasjon og rammeverk.

Det forutsettes at valg av personidentifikator skjer med en bred administrativ og politisk forankring som støtter opp om prosjektet. Det anbefales at det endelige valget av personidentifikator gjøres "irreversibelt" i den grad dette er konstitusjonelt mulig.

2 Innledning

2.1 Om Folkeregisteret

Bruken av Det sentrale folkeregister (Heretter: Folkeregisteret) har økt i takt med samfunnsutviklingen¹. I 1905 ble den fakultative (frivillige) folkeregistreringen innført i Norge. I 1915 kom den første loven som inneholdt bestemmelser om kommunens rett til å kreve medvirkning fra publikum ved opprettelse av register som påla publikum å melde flyttinger innen kommunen og mellom kommuner (1).

Først under okkupasjonstiden ble det ved "lov" innført plikt for alle landets kommuner til å innføre folkeregister. Samtidig ble den offentlige registrering av fødsler, vigsler og dødsfall overført til Folkeregisteret fra geistligheten. Stortinget vedtok "Lov om folkeregistre" i 1946. Opplysningene i registeret er basert på folketellingen avholdt samme år.

Siden har det skjedd betydelige endringer i Folkeregisterets oppgaver og samfunnets avhengighet til registeret. Dagens fødselsnummersystem ble etablert 1. oktober 1964 da oppbyggingen av et sentralt personregister startet. Alle som var bosatt i Norge på folketellingstidspunktet i 1960, fikk tildelt et eget nummer. Den eldste som fikk tildelt fødselsnummer, var født i 1855. Alle bosatte etter 1964 har fortløpende fått tildelt fødselsnummer (2). Fødselsnummer tildeles i dag de som er bosatt i Norge i folkeregisterlovens forstand, de som er født i Norge samt norske statsborgere som aldri har bodd i Norge men som skal ha norsk pass. Folkeregisteret inneholder i dag informasjon om over 7 millioner personer som er eller har vært bosatt i Norge.

Norge var sammen med de skandinaviske landene pioner på området, og fremdeles har mange land ikke et felles nasjonalt nummer eller annen felles personidentifikator for identifisering av landets innbyggere. Noen land har ulike systemer med sektorvise identifikatorer.

I 1970 ble det vedtatt ny lov om folkeregistrering, som fremdeles gjelder. Folkeregisteret er i dag organisert som en del av Skattedirektoratet. Det er et offentlig register hovedsakelig over personer som bor eller har vært bosatt i Norge. I tillegg til registrering av opplysninger på grunnlag av meldinger, har Folkeregisteret ansvar for blant annet prøving av ekteskapsvilkår, navneendringer, utstedelse av attester, fødselsnummer og D-nummertildeling (se 3.2 for omtale av D-nummer).

Folkeregisterets rolle har utviklet seg over tid og har blitt stadig viktigere. Opplysningene benyttes av både offentlige og private virksomheter. Det er krav om entydig identifisering av personer og at tidligere tildelte fødselsnumre ikke skal brukes om igjen, jfr. kapittel 3.4.

Registeret er sentralt i offentlig personrelatert saksbehandling og planlegging. Også for store deler av privat sektor er Folkeregisteret en nødvendig informasjonskilde for å utføre lovpålagte oppgaver.

Med utvikling av IKT og elektronisk kommunikasjon har et korrekt og oppdatert folkeregister bidratt til store effektivitetsgevinster for hele det norske samfunn. Folkeregisteret har blitt en nødvendig del av vår felles infrastruktur.

¹ Norge hadde kommunevise folkeregistre fram til 1992, da folkeregisteret ble digitalisert og landsdekkende, og ble til DSF.

2.2 Problemstilling

Med dagens populasjon har fødselsnummeret i Folkeregisteret grunnet kapasitetshensyn en antatt maksimal levetid frem t.o.m. 2039, ref. kapittel 3.1. På denne bakgrunn er det behov for å endre dagens identifikator. Finansdepartementet har gitt Skattedirektoratet i oppdrag å utrede nærmere de samfunnsmessige, økonomiske og administrative konsekvenser knyttet til valg av ny identifikator. Prosjektet sees i sammenheng med strategi for nytt folkeregister og inngår i Skattedirektoratets *Program for modernisering av Folkeregisteret*.

2.3 Mandat

Målene for utredningen er beskrevet slik i mandatet (utdrag):

Hovedmål

De samfunnsøkonomiske konsekvensene knyttet til ny identifikator utredes i et slikt omfang at det kan benyttes som et forsvarlig grunnlag for å treffe en beslutning om hvilket alternativ man bør velge i et fremtidig folkeregister.

Delmål

Utredningen skal oppfylle de krav og føringer som er stilt til konseptvalgutredninger i Finansdepartementets rammeverk for kvalitetssikring. Dette innebærer at analysen er inndelt i seks hovedkapitler, som gjenspeiler delmålene til prosjektet. Disse er:

- *Behovsanalyse*
- *Strategikapittel*
- *Overordnede krav*
- *Mulighetsstudie*
- *Alternativanalyse*
- *Føringer for forprosjektfasen*

Relevante problemstillinger

Det skal beskrives ulike overgangsstrategier. Overgang til ny identifikator bør kunne varsles i meget god tid, slik at tilpasninger i all hovedsak kan tas i forbindelse med ordinære oppgraderinger / utbyttinger av IT-systemer

Til grunn for den samfunnsøkonomiske analysen og estimeringen skal det ligge representative tall for de berørte virksomheter, jf. interessentanalysen. Tallene vil bli innhentet gjennom intervjuer eller andre hensiktsmessige metoder, for eksempel skjema. Datagrunnlaget må ha en så høy kvalitet at utredningens konklusjoner blir troverdig. Dette skal belyses i usikkerhetsanalysen.

2.4 Avgrensning og rammer for utredningen

Mandat angir avgrensning og rammer slik:

Utgangspunktet for utredningen er rapporten om utveksling av grunndata på personinformasjonsområdet fra 2007 samt resultater fra strategiarbeidet i forbindelse med Skattedirektoratets program Modernisering av Folkeregisteret. Brev fra Finansdepartementet, datert 19.11.2009, gir føringer for prosjektets resultat og gjennomføring. Utredningsinstruksen legges til grunn for gjennomføringen av utredningen.

I rapporten "Utteksling av grunndata på personinformasjonsområdet" er det foreslått to alternativer for ny identifikator. Primært foreslås en såkalt informasjonsløs identifikator, mens det sekundært foreslås at dagens fødselsnummer videreføres, men at dets oppbygning endres ved å endre et av sifrene fra å være kontrollsiffer til å inngå i løpenummeret for å øke kapasiteten. Disse to alternativene skal tas med i mulighetsstudien og alternativanalysen (1).

Det kan også være aktuelt å beskrive varianter av disse alternativene dersom det avdekkes at disse er potensielt bedre. Prosjektet skal undersøke og vurdere om det er konseptuelt andre alternativer som skal inngå i utredningen. Her bør løsninger valgt i andre sammenlignbare land inngå.

Det skal legges til grunn at ny personidentifikator skal kunne benyttes minst fram til år 2150. Videre skal det tas hensyn til mulige endringer i populasjonen slik det er anbefalt og eventuelt blir vedtatt i forbindelse med strategiarbeidet for modernisering av Folkeregisteret.

Finansdepartementet og Skattedirektoratet har bestemt at det gjennomføres en konseptvalgutredning (KVU) i henhold til Finansdepartementets retningslinjer og rammeverk. Hovedmålet for prosjektet er å utrede de samfunnsøkonomiske konsekvensene knyttet til ny personidentifikator i et slikt omfang at det kan benyttes som beslutningsgrunnlag for valg av alternativ.

Det gjennomføres derfor en samfunnsøkonomisk analyse med relevante alternative personidentifikatorer. De samlede nytte- og kostnadsvirkninger skal identifiseres, beskrives og kvantifiseres så langt det er forsvarlig.

2.5 Grensesnitt og føringer

Prosjektet er en del av Skattedirektoratets *Program for modernisering av Folkeregisteret* og skal sees i sammenheng med dette. I perioden januar til juni 2011 pågikk en bred strategiprosess med 28 viktige brukere og interessenter av Folkeregisteret. Dette arbeidet har resultert i utredningen "Modernisering av Folkeregisteret – rapport fra strategigruppen", som er en oppfølging av Grunndatarapporten. Rapporten ble levert til Finansdepartementet i november 2011. Temaet ny personidentifikator lå utenfor mandatet til denne rapporten.

Med utgangspunkt i dette arbeidet ble det våren 2012 satt i gang arbeid med forstudie for å konkretisere alternative modeller for et modernisert folkeregister, inklusiv samfunnsøkonomiske konsekvenser (3). Anbefalingene fra forstudien vil kunne påvirke vurderinger av alternative nye personidentifikatorer. Mest relevant er en mulig utvidelse av populasjonen som tildeles personnummer i framtiden. Det er aktuelt å fjerne dagens D-nummer og at denne gruppen i framtiden får fødselsnummer. Dette er et tema som adresseres til forprosjektrapporten, men som gir føringer for en ny personidentifikator. I denne rapporten er temaet behandlet i en egen robusthetsanalyse om kapasitet 7.2.2 Robusthet for utvidelse av dagens populasjon.

Høsten 2012 ble utredning av "Ny distribusjons- og prismodell for Folkeregisteret" avsluttet og oversendt Finansdepartementet for videre behandling (4). Denne utredningen har felles

berøringspunkter med denne KVUen, og prosjektene blir sett i sammenheng der dette er naturlig.

I april 2012 presenterte Statsministeren og Fornyings-, administrasjons- og kirkeministeren Regjeringens digitaliseringsprogram “På nett med innbyggerne”. I sammendraget heter det bl.a.:

“Regjeringen har store ambisjoner for offentlig sektor. Vi trenger en sterk og effektiv forvaltning for å sikre en god samfunnsutvikling i Norge. Digitalisering vil bidra til merkbare forbedringer på tvers av offentlig sektor i de kommende årene. Digitalisering vil både føre til et bedre og raskere møte med offentlig sektor for innbyggere og næringsliv, og bedre ressursbruk i offentlig sektor.

For å realisere fremtidens digitale forvaltning må visse forutsetninger være på plass. Det trengs en digital infrastruktur for forvaltningen. Infrastrukturen skal bestå av noen felleskomponenter som ulike deler av forvaltningen trenger, slik som elektronisk ID, digital postkasse og offentlige registre som understøtter den digitale forvaltningen...”

I kap 3.6 *Felleskomponentene skal ivareta offentlig sektors samlede behov* (s. 27) heter det:

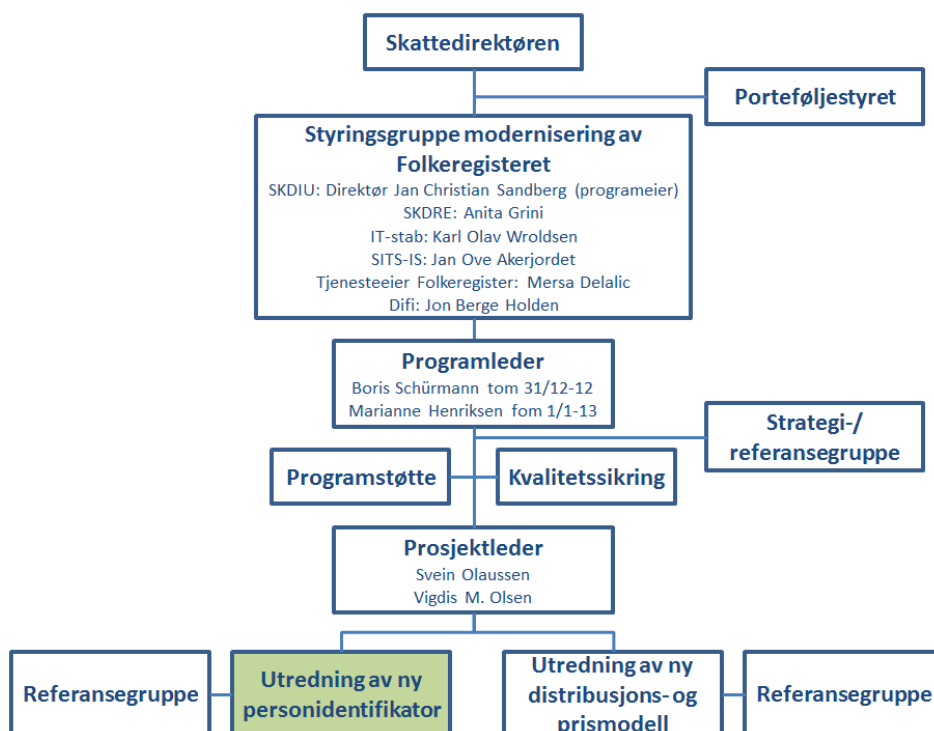
“Regjeringen vil bedre samordningen av departementenes arbeid med IKT-utvikling i forvaltningen. Altinn-komponentene, ID-porten, folkeregisteret, matrikkelen og enhetsregisteret utgjør sentrale elementer i den felles digitale infrastrukturen og er viktige for digitaliseringen av hele offentlig sektor”

2.6 Organisering

Utredningen av ny personidentifikator har vært et prosjekt under det pågående *Program for Modernisering av Folkeregisteret*.

I forbindelse med utredningen er det opprettet en referansegruppe med viktige interessenter fra offentlig og privat sektor. Referansegruppen har hatt 5 møter og gitt faglige innspill til arbeidet.

Nedenfor er tatt inn en figur som viser organiseringen av arbeidet:



Figur 1 Organisering av utredningsarbeidet

2.7 Utforming av KVUen

Mandatet angir at det skal utarbeides en KVU i henhold til de krav og føringer som er stilt til konseptvalgutredninger i Finansdepartementets rammeverk for kvalitetssikring.

Formålet med en KVU er å utarbeide et beslutningsgrunnlag for valg av hvilket konsept som eventuelt skal videreføres i en forprosjektfase.

I Finansdepartementets veileder nr. 11 Konseptvalg og detaljeringsgrad gis det signal om at det ikke er ønskelig med for høy detaljeringsgrad. Hvert konsept må imidlertid beskrives til den detaljeringsgraden som gjør det er mulig å bedømme det i forhold til andre konkurrerende prinsipielle løsningsforslag (5). Faktagrunnlaget som kan være mangelfullt, usikkert og med betydelige avvik, må suppleres med mer eller mindre sikre antakelser og skjønn.

Finansdepartementets rammeverk stiller krav om at en KVU inneholder seks hovedkapitler (behovsanalyse, strategikapittel, overordnet kravdokument, mulighetsstudie, alternativanalyse og føringer for forprosjektfasen) som utgjør en logisk sekvens.

KVUen er søkt utarbeidet i samsvar med Finansdepartementets rammeverk. Det er bl.a. søkt å sikre en direkte logisk årsakssammenheng mellom behovet som ligger til grunn for tiltaket, ref. avsnitt 2.2 og kapittel 3, de definerte målene og effekten som forventes oppnådd.

2.8 Referansedokumenter

Vedlagt følger bibliografi som viser viktige referansedokumenter som er benyttet i forbindelse med utredningen.

3 Behovsanalyse

Behovsanalysen beskriver eksisterende fødselsnummer (dagens situasjon), andre personnumre i bruk og personidentifikatorer i andre land, kartlegger interessenter og drøfter hva som er sentrale konseptuelle problemstillinger. Deretter analyseres og struktureres de ulike behov. Det prosjektutløsende behov defineres som den avgjørende begrunnelse for at dagens personidentifikator må erstattes. Avslutningsvis er behovene oppsummert og prioritert.

Behovene knyttet til større offentlige investeringsprosjekter grupperes gjerne slik:

- Normative behov: Behov som er nedfelt i overordnede politisk vedtatte, nasjonale målsettinger og lovverk, etc.
- Etterspørselsorienterte behov: Behov som følger av etterspørsel etter en bestemt tjeneste.
- Interessegroupebaserte behov: Interessegrouppens behov i forbindelse med et problemkompleks.

De beskrevne behov danner utgangspunkt for innholdet i kapitlene om strategi, mål og krav.

3.1 Dagens fødselsnummer

Som nevnt innledningsvis i kapittel 2.1, startet etableringen av dagens Folkeregister i 1964 og alle som var registrert i folketellingen i 1960 fikk tildelt fødselsnummer. Den tidligst fødte personen i registeret er født i 1855.

I dette avsnittet beskrives dagens fødselsnummer sin forankring i lov/forskrifter, oppbygning, bruk av kontrollsifre og kapasitet.

Forankring i lov og forskrifter

I Lov om folkeregistrering (Folkeregisterloven) som trådte i kraft februar 1970, står det bl.a. følgende i kapittel 1 Organisasjon:

§ 1. Det skal være ett sentralt folkeregister for Norge. I Det sentrale folkeregister registreres alle personer som:

- a) er eller har vært bosatt i Norge,*
- b) er født i Norge,*
- c) har fått tildelt fødselsnummer eller D-nummer*

§ 2. Staten sørger for folkeregistreringen og bærer kostnadene ved denne virksomheten.

§ 3. Skattekontoret er folkeregistertyndighet i første instans. Departementet fordeler oppgaver mellom skattekontorene. Skattedirektoratet er sentral folkeregistertyndighet.

§ 4. For enhver som er bosatt i Norge, fastsettes et fødselsnummer. For andre personer kan det fastsettes enten et fødselsnummer eller et D-nummer når det foreligger et begrunnet behov for det. Departementet gir nærmere regler om tildeling og endring av fødselsnummer og D-nummer.

I Forskrift om folkeregistrering (FOR-2007-11-09-1268) står det bl.a. følgende (utdrag):

§ 2-2 Fødselsnummer:

Fødselsnummeret skal ha elleve siffer. De seks første siffer består av vedkommendes fødselsdato i rekkefølge to siffer for dag, to for måned, to for år. De fem siste siffer, personnummeret, består av tre individualsiffer og to kontrollsiffer.

Som det fremgår av teksten i lov og forskrift er det en forutsetning i dagens praksis at hvert enkelt fødselsnummer og D-nummer skal være unike, se videre i kapittel 3.4.

I Folkeregisteret skal registreres alle personer som er eller har vært bosatt i Norge, er født i Norge eller har fått tildelt fødselsnummer eller D-nummer. I Forskrift om folkeregistrering § 4-1 er hovedregelen om å være bosatt formulert slik:

Personer som oppholder seg i norske kommuner i minst seks måneder regnes som bosatt i Norge, med mindre annet følger av bestemmelsene i § 4-2 til § 4-7.

Oppbygning

I Forskrift om folkeregistrering fra 2007 er som nevnt fødselsnummeret bestemt til å være på 11 siffer og med denne oppbyggingen:

$$D_1 D_2 M_1 M_2 \text{Å}_1 \text{Å}_2 I_1 I_2 I_3 K_1 K_2$$

der D: dato, M: måned, Å: år, I: individualsiffer/personnummer, K: kontrollsiffer, I_3 viser kvinne (like tall) eller mann (ulike tall).

Første kontrollsiffer beregnes av de ni første sifrene i fødselsnummeret, mens det i utregningen av andre kontrollsiffer inkluderes også det første kontrollsifferet. For nærmere beskrivelse for beregningen av kontrollsymboler i dagens fødselsnummer (se vedlegg 1, s. 4-5 og vedlegg 2).

Dagens fødselsnummer gir som det fremgår av definisjonen over både informasjon om fødselsdato og kjønn.

I Forskrift om folkeregistrering fra 2007 § 2-2 er forskjellige serier av individualsiffer definert for ulike tidsperioder. Gjennom en koding av første siffer i individnummeret fremgår det hvilket århundre en person er født. En bearbeidet oppstillingen av sammenhengen mellom individnummer og fødselsår er vist i tabellen nedenfor:

Første siffer ↓	To siste siffer i årstall		
	00-39	40-54	55-99
0	2000-2039	1940-1954	1955-1999
1	1900-1939	1940-1954	1955-1999
2			
3			
4			
5	2000-2039	2040-2054	1855-1899
6			Ikke i bruk
7			
8		1940-1954 (Reserve)	1955-1999 (Reserve)
9		2040-2054 (Hvis ikke brukt)	Ikke i bruk

Tabell 1: Koding av århundre. Sammenhengen første siffer i individnummer og fødselsår (6).

Som tabellen viser ble ikke 0-serien benyttet før i 1940. Den er derfor nå tilgjengelig fram til og med 2039.

Den såkalte “reserven” innbefatter serien 999-900² og skal dekke tidsperioden fra 1940 til 1999 ved tildeling av fødselsnummer i forbindelse med innvandring av personer født mellom 1949 og 1999. Per september 2012 er denne serien ikke påbegynt, men det er for enkelte datoer svært få ledige individnumre. Sannsynlighet er imidlertid liten for at reserveserien vil bli benyttet for tidsperioden før 1954, da innvandrere i hovedsak er personer under 50 - 60 år.

For tidsperioden 1855-1899 er benyttet serien 749-500.

Bruk av kontrollcifre

Feilkorrigerende (eller feildetekterende) koder brukes til å korrigere (oppdage) feil ved overføring eller lagring av data. Idéen er å tilføre redundans³ i form av kontrollcifre til dataene slik at dersom feil oppstår kan vi oppdage og eventuelt korrigere feilene.

Generelt kreves det flere kontrollcifre for å korrigere en feil enn for bare å oppdage den. I dagens personnummersystem brukes to kontrollcifre som oppdager flere typer av feil. Dette inkluderer enhver feil i et enkelt siffer og alle ombyttinger av data i naboposisjoner og en del andre typer feil som ombytting av måned og dato.

Dersom kontrollsymbolene velges riktig oppdages flere feil. Ulempen er at kontrollcifre øker redundansen og ikke øker informasjonskapasiteten. I de fleste nasjonale personregistersystem brukes 0, 1 eller 2 kontrollcifre.

Prinsipper for feilkontroll er beskrevet i rapporten til Selmersenteret (vedlegg 1).

Kapasitet

Ethvert system for informasjonsbærende personidentifikatorer med fast lengde og alfabet vil ha et endelig antall unike identifikatorverdier per dato. Begrepet kapasitet uttrykker, innenfor konteksten av denne rapporten, en numerisk beskrivelse av hvor mange forskjellige identifikatorverdier som kan tildeles for et gitt valg av identifikatorsystem. De forskjellige identifikatorsystemene er forskjellige i bruk, og det er derfor hensiktsmessig å bruke forskjellige eksakte definisjoner i hvert tilfelle (vedlegg 3).

For identifikatorer som inneholder eksplisitt informasjon, er det hensiktsmessig å uttrykke kapasitet med hensyn på hver spesifikk verdi av det eksplisitte informasjonsinnholdet. For eksempel kan man uttrykke kapasitet for dagens fødselsnummerordning i forhold til antall forskjellige tilgjengelige identifikatorverdier for hver verdi av “DDMMÅÅ, kjønn, århundre”. I praksis vil man gjerne anta at det for hver fødselsdag i gjennomsnitt finnes like mange menn og kvinner, men kjønnsdeling av populasjonen vil likevel øke den relative variansen innenfor hver av de to gruppene sammenlignet med hele populasjonen. Ut fra denne forutsetningen er det rimelig å definere kapasitet for dagens fødselsnummer som antall tilgjengelige løpenummerverdier for hver dato.

For en informasjonsløs identifikator er det i prinsippet ingen individinformasjon innbakt i identifikatorverdien tilordnet hvert individ. Identifikatorverdier kan derfor tildeles fortløpende ut fra den totale mengden av «lovlige» identifikatorverdier. Dette gir en effektiv utnyttelse av

² Det telles fra 999 og nedover

³ Wikipedia: Redundans er informasjon som gjentar allerede etablert kunnskap uten å tilføre noe nytt

mengden av lovlige verdier, og det gir mening å definere kapasiteten ganske enkelt som det totale antallet forskjellige identifikatorverdier.

For dagens fødselsnummer kan ikke alle individnumre benyttes da enkelte ikke godkjennes av moduluskontrollen som er bygd inn for å avsløre eventuelle feil i fødselsnummeret. I gjennomsnitt vil vi i 1 av 11 tilfeller få verdien $K_1 = 10$, noe som betyr at det respektive løpenummeret ($I_1 I_2 I_3$) må forkastes da kun sifrene 0 til 9 kan benyttes for K_1 . Tilsvarende gjelder for K_2 . Av en serie på 100 numre forkastes derfor ca. 17 individnumre, beregningen blir $10/11 \cdot 10/11 = 100/121 = 82,6 \%$. Dermed er det bare 826 tilgjengelige individnumre i en serie på 1000 ($I_1 I_2 I_3$). Nærmere forklaring om bruk av kontrollcifre, moduluskontroll og kapasitet er gitt i rapporten fra Seltersenteret (vedlegg 1), se spesielt kapittel 3.3, 4.1, 4.3.

Nedenfor er det gjort en oppstilling av tilgjengelig kapasitet av individnumre for dagens fødselsnummer for de ulike tidsperioder:

	Første siffer i individualsiffer										Kapasitet per dag
	0	1	2	3	4	5	6	7	8	9	
1855-1899	-	-	-	-	-	83	83	41	-	-	207
1900-1939	-	83	83	83	83	-	-	-	-	-	331
1940-1999 (eks. reserve)	83	83	83	83	83	-	-	-	-	-	413
1940-1999 (inkl. reserven)	83	83	83	83	83	-	-	-	-	83	496
2000-2039	83	-	-	-	-	83	83	83	83	83	496
2040-2054 (eks. reserve)	-	-	-	-	-	83	83	83	83	-	331
2040-2054 (inkl. reserve)	-	-	-	-	-	83	83	83	83	83	413

Tabell 2: Antall tilgjengelige individnumre per dag for tidsperioder (s. 10 i rapport fra Seltersenteret (6)).

I perioden 2000-2039 vil kapasiteten være 496 individnumre. Her kan serien 099-000 også benyttes. Reserveserien 999-900 er kun aktuell for fødte i perioden 1940-1999 og vil dermed tidligst kunne "kollidere" med numre fra 20 århundre i 2040. Serien 999-900 kan dermed også benyttes.

Med stor sannsynlighet vil ikke reserveserien 999-900 for perioden 1940-1954 (1999) bli brukt⁴. Det betyr at systemet for fødselsnummer har en kapasitet på 413 individnumre per dato for årene 2040-2054. Det kan her være et scenario der reserveserien er benyttet, men kun for år etter 1954. Serien 999-000 kan dermed frigjøres for perioden fra 2040 til og med 2054.

Hvis derimot reserveserien tas i bruk for tidsperioden 1940-1954, dvs. tildeles personer som kommer til Norge gjennom innvandring med fødselsdato mellom 1940 og 1954, vil kapasiteten fra og med 2040 kunne bli redusert ned mot 330 individnumre per dato.

Fra og med 2055 vil kapasiteten bli svært redusert, ref. Tabell 1. Hvis reserveserien fortsatt er tilgjengelig, vil kapasiteten kunne bli 207 ($83+83+41$) individnumre per dato. I et scenario der reserveserien er benyttet, vil kapasiteten kunne bli redusert helt ned mot 124 individnumre per dato.

3.2 Andre personnumre i bruk

I Folkeregisteret føres det i dag både fødselsnummer og D-nummer. I tillegg har Utlendingsdirektoratet (UDI) etablert et eget system med DUF-nummer for utenlandske borgere

⁴ Det er lite trolig at reserveserien blir tatt i bruk da det for disse årene (perioden 1940-1954) ikke er noen datoer som er kritiske, og innvandring av personer fylt 60 år er lav.

som er i kontakt med UDI og som ikke får rekvirert D-nummer. Personer med DUF-nummer registreres i en egen database hos UDI (utlendingsdatabasen eller UDB). Dessuten har politiet og helsevesenet sine egne personidentifikatorer.

Nedenfor er gitt beskrivelser av D-nummer, DUF-nummer og andre personidentifikatorer.

D-nummeret

Ordningen med D-nummer ble tatt i bruk i 1978, opprinnelig for utenlandske sjømenn på norsk-registrerte fartøy i utenrikshandel. I Forskrift om folkeregistrering fra 2007 er det i § 2-5 gitt følgende beskrivelse av D-nummer:

D-nummeret skal ha elleve siffer. De seks første siffer består av vedkommendes fødselsdato i rekkefølge to siffer for dag der første er tillagt fire, to for måned, to for år. De fem siste siffer, personnummeret, består av tre individsiffer og to kontrollsiffer.

Som individsiffer for personer født:

- a) før 1. januar 2000 brukes serien 000-499,
- b) etter 1. januar 2000 brukes serien 500-999.

Kvinner skal ha like tall og menn ulike tall som siste individsiffer.

D-nummeret kan bare endres når fødselsdato eller kjønnsstatus endres.

D-nummeret tildeles i Det sentrale folkeregister og administreres av Skattedirektoratet. Nærmere bestemmelser om tildelingsmyndighet gis av Skattedirektoratet.

Vi ser at D-nummeret er konstruert på samme måte som fødselsnummeret bortsett fra at:

- På første siffer for dag er det lagt til 4
- Seriene for å skille århundre er noe annerledes (med samme hovedtrekk)

I Forskrift om folkeregistrering fra 2007, § 2-6 er det angitt hvem som kan tildeles D-nummer:

D-nummer tildeles fysisk person som ikke fyller vilkåret for å få tildelt fødselsnummer, og som er:

- a) skatte- eller avgiftspliktig til Norge, herunder til Svalbard, jf. Svalbardskatteloven § 2-1 og § 2-2
- b) i forretningsforhold med norsk bank eller annen institusjon som er underlagt reglene i lov 10. juni 1988 nr. 40 om finansieringsvirksomhet og finansinstitusjoner (finansieringsvirksomhetsloven),
- c) registreringspliktig i Foretaksregisteret, Løsøreregisteret eller Konkursregisteret,
- d) rolleindehaver i juridiske enheter, jf. Enhetsregisterloven § 5 annet ledd bokstav f og § 6 første ledd bokstav a-e og h, eller har tilsvarende rolle i utenlandsk deltakerlignet selskap (DLS) eller i selskap som nevnt i skatteloven § 2-4 eller som utfører rapporteringsoppgaver på vegne av slike rolleindehavere,
- e) kontohaver i Verdipapirsentralen,
- f) eier av fast eiendom som omfattes av Matrikkellova,
- g) omfattet av ordning som forvaltes av Arbeids- og velferdsetaten eller Helseøkonomiforvaltningen, eller som har rettighet utledet fra slik person,
- h) bosatt på Svalbard, jf. forskrift om register over befolkningen på Svalbard § 2,
- i) under autorisering av Statens autorisasjonskontor for helsepersonell (SAFH)

Som listen viser vil dette kunne innbefatte en rekke personer med ulik tilknytning til Norge.

Det er to viktige forskjeller mellom fødselsnummeret og D-nummeret:

- Det registreres flere opplysninger om personer med fødselsnummer enn personer kun med D-nummer.
- Det er en mer omfattende kvalitetssikring ved rekvirering av nye nummer og bedre rutiner for oppdatering av informasjon rundt fødselsnummeret enn for D-nummeret.

Totalt er det i dag registrert ca. 1,3 mill. D-numre i Folkeregisteret.

DUF-nummer

DUF-nummeret tildeles alle asylsøkere og flyktninger som kommer til Norge. Dette er deres registreringsnummer i Datasystemet for utenlands- og flyktningesaker hos UDI også kalt utlendingsdatabasen eller UDB. Alle innvandrere som ikke kommer fra et EØS-land skal i utgangspunktet registreres i dette registeret når de ankommer Norge.

DUF-nummeret har tolv siffer og inneholder ikke informasjon om fødselsdato. Dersom det gis opphold i Norge, og personen blir her i mer enn seks måneder, blir det registrert fødselsnummer i Folkeregisteret. Da blir også DUF-nummeret registrert i Folkeregisteret.

DUF-nummeret kan betegnes som UDIs kundennummer (saksnummer).

Personer med DUF-nummer er ikke skatte- eller trygdepliktige, og det er derfor i mindre grad knyttet problemer med trygdemisbruk til den gruppen som kun er registrert med DUF-nummer. Denne gruppen har dog lovfestet rett til helseundersøkelser og helsehjelp (primær/spesialist), kommunale tjenester som skolegang (barnekonvensjonen) etc..

Andre personidentifikatorer

Det er registrert bruk av annen personidentifikator både i helsevesenet og politiet.

I helsesektoren benyttes hjelpenummer når en persons fødselsnummer eller D-nummer ikke er kjent. Person- og pasientopplysninger knyttes til hjelpenummeret. Det er definert et "felles hjelpenummer" som er unikt i hele sektoren, det er et informasjonsløst 11-sifret nummer som begynner med tallet 8 eller 9 og har kontrollrutine lik fødselsnummeret.

I politiets datasystemer benyttes ordinært fødselsnummer og en generert identifikator for personer som ikke har fødselsnummer. Det genererte fødselsnummeret er utformet som et vanlig fødselsnummer på elleve siffer, men hvor det legges til 20 på måned. Bruken av genererte nummer har økt de siste årene.

3.3 Andre land

Seltersenteret beskriver noen løsninger i andre land i vedlegg 1, appendiks 4 (7). Videre ble det gjort en vurdering av andre lands løsninger som inndata og grunnlag for mulighetsstudien.

3.4 Normative behov

Med normative behov menes lover og forskrifter, overordnede politiske mål, gitte rammebetingelser etc. Disse behovene skal ha et nasjonalt perspektiv og kan bli førende for tiltaket.

Det er her fokus på samfunnsbehov i lys av underoppfyllelse av "nasjonale" behov gjerne konkretisert i gjeldende lovverk, politiske vedtak eller målsettinger vedtatt av Storting etc.

I kapitlene 2.1 og 3.1 er de gjeldende lovmessige forankringer til Folkeregisteret beskrevet. Hovedpunktene her var:

- Det skal være ett sentralt folkeregister for Norge.
- For enhver som er bosatt i Norge, fastsettes et fødselsnummer.

Siste prikkpunkt lyder slik i loven:

“§ 4. For enhver som er bosatt i Norge, fastsettes et fødselsnummer. For andre personer kan det fastsettes enten et fødselsnummer eller et D-nummer når det foreligger et begrunnet behov for det.”

I Ot.prp. nr. 10 (1969-70) beskrives fødselsnummersystemet som et identitetsnummersystem for alle som er bosatt i riket. Under kommentarene til § 4 heter det bl.a.:

“Fødselsnummeret er en viktig identitetsbetegnelse for den registrerte befolkning.”

Loven angir dermed at hver enkelt person som er registrert i Folkeregisteret skal ha et *unikt* fødselsnummer.

Fødselsnummeret er derfor definert som en personidentifikator. I rapporten til Selmersenteret defineres en personidentifikator som følger (vedlegg 1):

“En personidentifikator er en kort sekvens av tegn hentet fra et predefinert tegnsett eller alfabet, som skal brukes som et entydig ”navn” for å identifisere en person”.

Man kan si at fødselsnummeret er en enkel og lett tilgjengelig personidentifikator for offentlige myndigheter og andre for entydig å kunne identifisere en person. Da fødselsnummeret ble innført i 1964, skjedde dette i forbindelse med opprettelsen av Det sentrale personregister. Det var et behov for entydig å kunne identifisere skatteyttere i forbindelse med ligningsbehandling- og skatteoppkreving.

I Norge benyttes fødselsnummeret i dag til å utføre en rekke lovpålagte oppgaver. Det kan være i forbindelse med:

- Rapportering av lønns- og trekkoppgaver til skattemyndighetene,
- Rapportering fra bank- og finansinstitusjoner til skattemyndighetene,
- Saksbehandling og utbetaling av dagpenger, arbeidsavklaringspenger, sykepenger, pensjon, barnetrygd og kontantstøtte (NAV),
- Registrering av personopplysninger i helsevesenet,
- Utarbeiding av offisiell personstatistikk gjennom gjenbruk av administrative personopplysninger,
- Kommunale oppgaver som barnehager, skole, syke- og aldershjem etc.,
- Registrering og oppfølging av bøter og avgifter (politi, tollvesenet etc.).

I nevnte sammenhenger benyttes fødselsnummeret som en unik personidentifikator hvor opplysninger tilknyttes en bestemt person. SSB bruker fødselsnummer for kopling av informasjon fra mange administrative registre til statistikkformål.

I Regjeringens digitaliseringsprogram *“På nett med innbyggerne”* som ble presentert i april 2012, beskrives en satsing på digitalisering som både skal føre til et bedre og raskere møte med offentlig sektor for innbyggere og næringsliv, samt bedre ressursbruk i offentlig sektor. Visse forutsetninger må være på plass, slik som en digital infrastruktur bestående av felleskomponenter, deriblant Folkeregisteret. I dette ligger en forventning om at vi i Norge skal ha et Folkeregister med høy kvalitet, som gir en entydig og komplett oversikt over landets innbyggere. Dette skjer gjennom bl.a. en unik personidentifikator med tilknyttet informasjon om den enkelte person.

Det forutsettes at en ny personidentifikator må følge relevante lover og regler for personvern.

3.5 Etterspørselsbaserte behov

Dagens fødselsnummer er konstruert og benyttet slik at den maksimalt kan opprettholdes for personer født i en tidsperiode på to hundre år. Imidlertid vil tilgjengelige nummerserier ha så liten kapasitet at siste år nåværende system kan benyttes med dagens populasjon er forventet å bli 2039, ref. kapittel 3.1 og avsnittet om kapasitet.

Etterspørsel i denne sammenheng er antall personer som er kvalifisert for registrering i Folkeregisteret. Innenfor dagens ordning og regelverk er etterspørselen eller behovet for fødselsnummer antall fødte per dato. Inkludert her er godkjent innvandring som omfatter personer som er bosatt i Norge, men er født i utlandet og har foreldre som begge er født i utlandet.

Kapasitet for dagens fødselsnummer er nærmere drøftet i kapittel 3.1, definert som antall tilgjengelige løpenummerverdier for hver dato. For informasjonsløse identifikatorer vil kapasiteten være det totale antall forskjellige verdier den enkelte identifikator kan ha. Når vi skal se på etterspørsel er det derfor naturlig å se både på tilgjengelige løpenummer per dato og det kumulative behovet.

Statistisk sentralbyrå har ikke utgitt befolkningsprognoser lenger frem enn til år 2100. For å finne behovsestimater for år 2150 er det gjort egne anslag basert på en enkel framskriving av Statistisk sentralbyrå sine prognoser fram til 2100.

Det er vanskelig å lage pålitelige prognoser for hvor mange personidentifikatorer som er nødvendig for en gitt dag eller periode da prognosen avhenger av usikre forhold som:

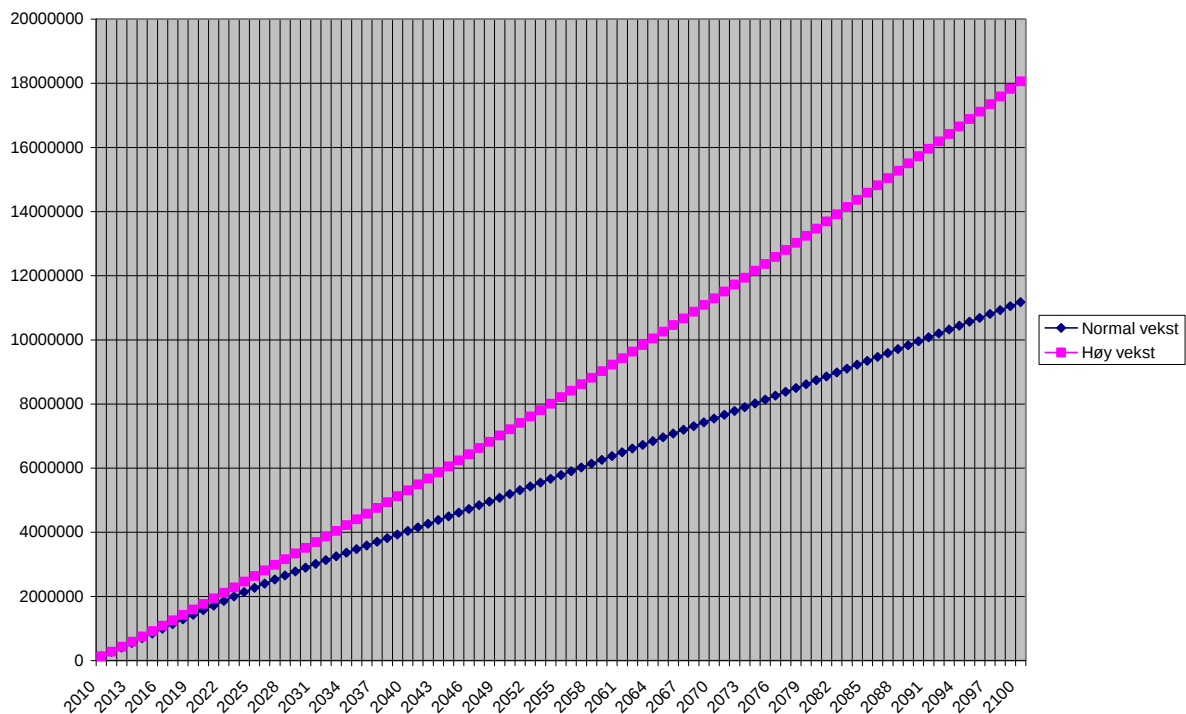
- Biologiske fødselsprosesser som avhenger av blant annet befolkningsstørrelse, alderssammensetning og enkelte endringsprosesser i samfunnet.
- Hvordan fødsler fordeler seg utover kalenderåret.
- Innvandring som er avhengig av regionale og globale økonomiske og politiske forhold.

Statistisk Sentralbyrå har utarbeidet prognoser for befolkningsutviklingen basert på forskjellige forutsetninger for fødsler og innvandring uttrykt gjennom lav, normal og høy vekst. I denne rapporten har vi fokusert på scenariene for normal og høy vekst da lav vekst scenariet er for risikabelt å legge til grunn for estimering av fremtidig behov.

Selmersenteret har i sin rapport illustrert ulike scenarier for etterspørsel av personidentifikator basert på nevnte statistikk fra Statistisk sentralbyrå (6). Avsnittene nedenfor er et utdrag fra appendiks 5 i Selmersenteret sin rapport (vedlegg 1).

Informasjonsløs personidentifikator

Det kumulative behovet som er relevant for en informasjonsløs personidentifikator, er vist nedenfor fram til 2100 basert på Statistisk sentralbyrå sine befolkningsprognoser med normal og høy vekstrate:



Figur 2 Kumulativt behov for nye personidentifikatorer, inkludert fødsler og innvandring

Figuren viser at:

- Behovet for nye identifikatorverdier ved høy vekstrate er ventet å passere 4 millioner i 2033, 8 millioner i 2054, 12 millioner i 2074, og 18 millioner i 2100
- Ved normal vekstrate er behovet noe over 11 millioner i 2100

En enkel framskriving av Statistisk sentralbyrå sine prognoser fram til 2150 indikerer behov for ca. 30 millioner og 18 millioner identifikatorverdier ved henholdsvis høy og normal vekstrate.

Dersom eksisterende befolkning også skal tildeles en ny informasjonsløs personidentifikator, øker behovet med ca. 5 millioner og over 7 millioner hvis også døde inkluderes. I tilfellet der alle som i dag er registrert i Folkeregisteret også skal tildeles ny identifikator, er behovet på respektive 37 og 25 millioner identifikatorverdier fram til 2150.

Alle prognoser og anslag er her beheftet med stor usikkerhet. Av den grunn kan det være fornuftig å legge høyt anslag til grunn når kapasitetsbehovet skal angis for tidsrommet fram til 2150. Med et slikt resonnement til grunn må en ny informasjonsløs personidentifikator ha kapasitet på minst 37 millioner identifikatorverdier, dette blir et krav til kapasitet.

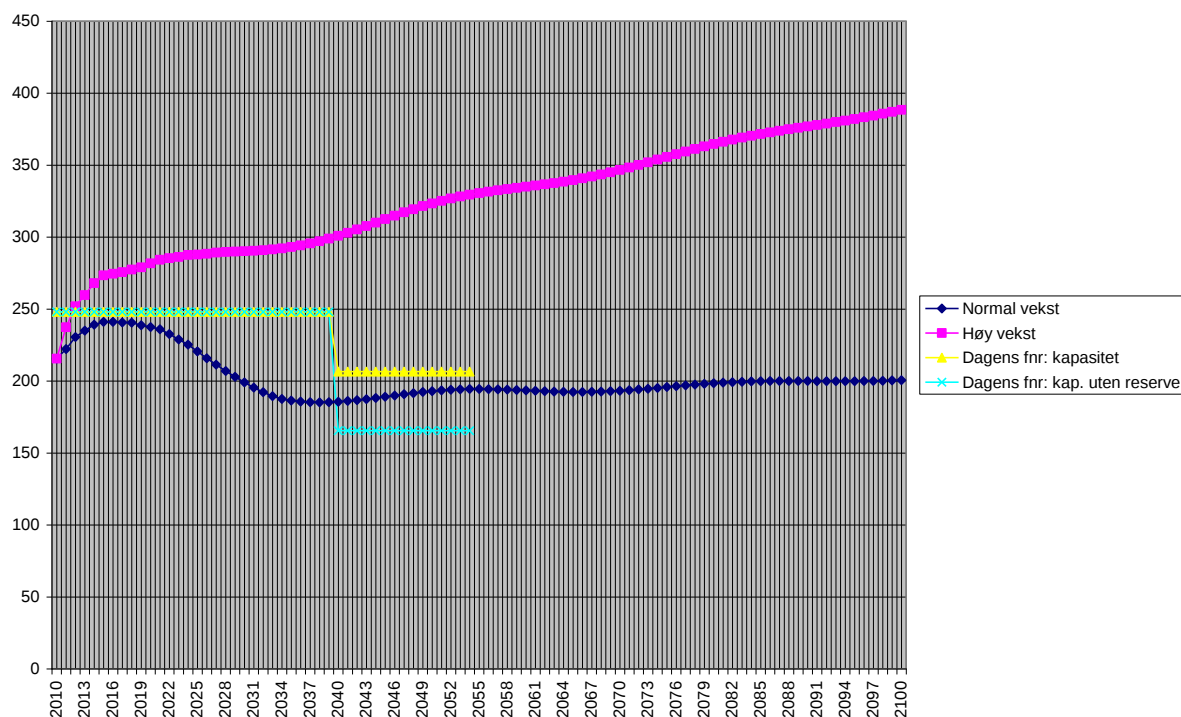
Informasjonsbærende personidentifikator

For en personidentifikator som inkluderer informasjon om fødselsdag, er det kritiske kapasitetsbehovet bestemt ved hvor mange som blir født hver enkelt dag i gyldig populasjon. Innvandrere vil få tildelt personidentifikator basert på fødselsdato og kjønn når de blir registrert i Folkeregisteret.

Antall fødsler er normalt ikke jevnt fordelt over årets dager da fødsler til en viss grad er "sesongbetont". Dessuten er det ikke vanlig å sette i gang fødsler eller å foreta planlagte keisersnitt i helger og høytider. Analysene fra Statistisk sentralbyrå tar utgangspunkt i at på årets "travleste" dag blir det født et antall individ som overstiger årsgjennomsnittet med 14 %.

Når vi deler populasjonen inn i kjønn, vil den relative variansen med hensyn til antall individer født per dag innenfor hvert kjønn være større enn for populasjonen som helhet. Vi antar at halveringen av gruppen (deling av nummerserien for kvinner og menn) medfører en økning av variansen med en faktor på 1.4 (kvadratroten av 2), slik at det på årets travleste dag kan bli født 20 % flere enn gjennomsnittet per kjønn.

Figuren nedenfor viser behovet for personidentifikatorer per kjønn på årets travleste dag når kjønn er den del av informasjonen:



Figur 3 Behov for nye personidentifikatorer per kjønn på dagen med flest fødsler per år inkludert innvandring

I figuren er kapasiteten i dagens fødselsnummersystem tegnet inn med og uten reserveserien (9-serien).

Figuren viser at behovet ved høy vekstrate vil overstige kapasiteten for dagens fødselsnummer allerede i 2013. Ved normalvekstrate vil kapasiteten ved dagens fødselsnummer akkurat være tilstrekkelig for toppårene 2015 – 2018 og er videre tilstrekkelig fram til og med 2039. Fra 2040 vil man med små marginer ha nok kapasitet hvis reserveserien er disponibel. Så langt fram i tid er prognosene svært usikre og marginene betraktes som for små til å kunne fastslå at kapasiteten er tilstrekkelig fra og med 2040. Hvis derimot reserveserien ikke er disponibel, forventes at behovet er større enn kapasiteten fra 2040. Dette underbygger mandatet der det er satt som forutsetning at ny personidentifikator må være tatt i bruk senest fra 2040.

Figuren viser videre at:

- Kapasitetsbehovet per dag ved høy vekst er ventet å passere 250 personidentifikatorer per kjønn i 2012, 300 i 2040, og 350 i 2072 og litt over 380 i 2100.
- Ved normal vekst er det ventet at kapasitetsbehovet per dag når en topp i underkant av 250 personidentifikatorer per kjønn rundt år 2016, og at det fra år 2028 vil stabilisere seg på verdier under og rundt 200 identifikatorer per dag.

Selv om det fødes noen flere gutter enn jenter, vil det være naturlig og mest praktisk å sette av like store nummerserier for både jenter og gutter. De oppgitte verdiene i figuren over kan dermed dobles for å angi det totale behovet. Det innebærer kapasitet på 770 og 400 personidentifikatorer i 2100 ved respektive høy og normal vekstrate.

En enkel framskriving av Statistisk sentralbyrå sine prognoser gir et behovsanslag på 900 og 420 personidentifikatorer per dato i 2150 ved henholdsvis høy og normal vekstrate.

Alle prognoser og anslag er her beheftet med stor usikkerhet. Av den grunn kan det være fornuftig å legge høyt anslag til grunn når kapasitetsbehovet skal angis for 2150. Med et slikt resonnement må en ny personidentifikator ha kapasitet på minst 900 ulike verdier (begge kjønn) per dato i 2150.

Endringer i populasjonen (kriterier for hvem som skal ha fødselsnummer)

En stor usikkerhet knyttet til fremtidig behov for personidentifikatorer er mulige endringer i populasjonen som skal registreres i Folkeregisteret. Mandatet for KVUen legger til grunn dagens lover og regler for definisjon av populasjonen.

Det pågående *Program for modernisering av Folkeregisteret* vurderer den fremtidige populasjonen. Hvis beslutningsprosessen her resulterer i at populasjonen skal endres, vil det kunne medføre endrede krav til kapasitet for ny personidentifikator.

Mandatet for KVUen angir at det tas hensyn til mulige endringer i populasjonen slik det er anbefalt og eventuelt blir vedtatt i forbindelse med strategiarbeidet for modernisering av Folkeregisteret. Strategiarbeidet er imidlertid fortsatt pågående, slik at i prosessen for denne KVUen er det derfor bestemt at mulige endringer i populasjonen skal vurderes gjennom en robusthetsanalyse. Dette er gjort i kapittel 7.2 i den samfunnsøkonomiske analysen.

Spesielle hendelser

Av og til kan det inntreffe store hendelser som får konsekvenser på et lands befolkningsutvikling. Det kan være naturkatastrofer, sykdom/epidemier, kriger etc.

Den verste katastrofen i Norge er Svartedauden som kom til Norge i 1349 og medførte at over halvparten av befolkningen døde. Mange pester har også satt sine dype spor i befolkningen. Etter Spanskesyken rundt 100 år tilbake har imidlertid Norge på mange måter vært forskånet fra slike hendelser som har påvirket befolkningsutviklingen i stor grad.

2. verdenskrig satte imidlertid sine synlige spor med lave fødselsrater. Frigjøringen våren 1945 medførte fødselstall i 1946 som ennå er de høyeste i landets historie.

I vårt moderne samfunn, der ofte begge foreldre er i arbeid og hvor forventninger og krav til velstand er store, har dette ført til relativt lave fødselsrater i et historisk perspektiv. Samtidig er innvandringen til Norge blitt høyere enn tidligere. Gjennomsnittlig antall fødte barn per kvinne har imidlertid vært ganske stabilt i Norge de siste 40 årene og vært noe under 2. Dette er høyere enn for mange sammenlignbare land.

Store hendelser som påvirker befolkningsutviklingen er vanskelig å forutse og dermed legge inn i behovsestimatene. Dette er derfor ikke forsøkt gjort i denne utredningen.

Utviklingen videre for fødselsrater og innvandring er vanskelig å spå, men skal i prinsippet være innebygget i Statistisk sentralbyrå sine prognoser. Imidlertid kan en trekke den konklusjonen at gode marginer over krav som er nødvendig for å dekke kjente forutsetninger og scenarier, bør tillegges vekt i den samfunnsøkonomiske analysen.

3.6 Interessentgruppebaserte behov

Det er gjennomført en interessentanalyse for å få kartlagt viktige interesser som berøres og hva som er aktørenes preferanser og opplevde behov i forhold til å innføre en ny personidentifikator i Norge.

Analysen omfatter ikke samtlige interessenter, men gir en oversikt over viktige interessenter som dette utredningsprosjektet har vært i kontakt med eller har informasjon om.

Interessentanalyse

I realiteten er alle bosatte i Norge og alle norske virksomheter brukere av dagens fødselsnummer, og vil bli påvirket i større eller mindre grad av at dagens identifikator endres. I hvilken grad de vil bli berørt, avhenger av hvilken ny identifikator med tilhørende overgangsordning som velges.

Vi har gruppert interessentene etter primære, sekundære og øvrige interessenter ut i fra følgende definisjoner:

Primære interessenter

- A. Eier av systemet
- B. Brukere som utøver myndighet og brukere med lovpålagte oppgaver som betinger bruk av systemet.

Sekundære interessenter

- A. Brukere som ikke utøver myndighet eller har lovpålagte oppgaver, men der valg av konsept har direkte betydning
- B. Kravsettere
- C. Personer registrert i Folkeregisteret

Øvrige interessenter

- A: Brukere der valg av konsept har mindre eller ingen betydning

Det er sendt ut spørreskjema i forbindelse med utredningen til en rekke primære og sekundære interessenter, for å kartlegge viktige behov og konsekvenser ved innføring av ny personidentifikator. I tillegg har det vært oppfølgingsmøter og intervju med de fleste av interessentene som deltok i undersøkelsen for å verifisere og utdype mottatt informasjon. Det har vist seg nyttig og nødvendig å gå i dybden og bruke tid for å kvalitetssikre informasjonen.

Kartleggingen er også basert på tilgjengelig informasjon, høringsuttalelser fra virksomheter til Grunndatarapporten (Finansdepartementet 2008) og spørreskjemaundersøkelse med utvalgte virksomheter i forbindelse med strategiprosessen til *Program for modernisering av Folkeregisteret* (Skattedirektoratet 2011).

Foruten å gruppere i primær, sekundær og øvrige interessenter, er interessentene delt inn i grupper etter virksomhetsområder som de naturlig tilhører. Denne grupperingen er videreført til den samfunnsøkonomiske analysen.

Tabellen under viser oversikt over kartlagte grupper av interessenter og deres overordnede hovedsyn.

Sektor	Virksomhet	Hovedsynspunkter
Statlige virksomheter med personopplysninger som kjerneoppgave	UDI	I høy grad påvirket av ny personidentifikator.
	Statens kartverk	Ønsker overgangsperiode på mellom 5-10 år.
	NAV	Behov for å koordinere prosjektgjennomføringen mellom virksomhetene, pga. gjensidig avhengighet.
	Toll- og avgiftsdirektoratet	Store konverteringskostnader.
	Skattedirektoratet	To parallelle nummerserier vil også være krevende i forbindelse med konvertering.
	Politiet/Kripos	
	Statistisk sentralbyrå	
	Brønnøysundregistrene	
Helsesektoren	Statens vegvesen	
	Primærhelsetjeneste	I stor grad påvirket. Stor og kompleks sektor. Stort antall systemer. Lang innføringstid, f.eks. 20 år. Ønsker beslutning av identifikator raskt. Konvertering medfører fare for liv og helse, og er ikke gjennomførbart for helsesektoren.
	Spesialisthelsetjeneste	
	Apotek og bandasjister (private)	
	Private spesialisthelsetjeneste	
	Private primærhelsetjenester	
Andre statlige virksomheter	Folkehelseinstituttet	
	Helsedirektoratet	
	Difi	Er påvirket av ny personidentifikator.
	Lånekassen	Noen ønsker overgangsperiode på 10 år, andre ønsker kortere periode (5 år).
	Posten	Lang overgangsperiode kan også være krevende, og vil øke forvaltningskostnadene.
	Uninett	
	Senter for IKT i utdanningen	
	Universitetet i Oslo	
	Forsvaret	
Kommuner og fylkeskommuner	Fylkesmannen	
	Kommuner og fylkeskommuner	I høy grad påvirket av ny personidentifikator. Ønsker overgangsperiode på 10 år. Usikkert om kommunesektoren vil kunne gjennomføre konvertering. Utfordring at kommunesektoren består av mange enheter.

Systemleverandører	EVRY, SAP, Unit4 Agresso, Visma, Aditro	I høy grad påvirket, men har stor grad av fleksibilitet i sine løsninger slik at konsekvensene er lave.
Bedrifter	Finansnæringen Telesektoren Andre bedrifter	Finansnæringen er i høy grad berørt av ny personidentifikator. Ønsker overgangsperiode på ca. 10 år. Gamle stormaskiner gir ekstra kostnader for deler av næringen. Det er dog store forskjeller mellom virksomhetene, og hvor store de økonomiske konsekvensene er estimert til. Konvertering gir svært høy risiko. Andre næringer er i mindre grad berørt.

Tabell 3: Innspill fra interessenter

Noen av hovedfunnene fra interessentanalysen er her beskrevet og utdypet nærmere. De er systematisert på temaer, med angivelse av de viktigste av interessentenes behov. Behov innbefatter her også ønsker om egenskaper og utforming av en ny personidentifikator.

Informasjonsløs personidentifikator

Interessenter som argumenterer for en informasjonsløs personidentifikator er hovedsakelig større offentlige virksomheter med stor kontaktflate mot utenlandske borgere i forbindelse med innvandring, andre former for opphold i Norge og stønader, ytelser etc. Dette er en befolkningsmasse med stor bevegelse og som derfor endrer seg over tid. En del av de personene som disse virksomhetene er i kontakt med, skal på et senere tidspunkt registreres i Folkeregisteret. Nevnte virksomheter har ofte egne felt for personers fødselsdato og kjønn i sine systemer, slik at de ikke er avhengig av fødselsnummeret for tilgang til disse opplysningene.

Mange av personene har mangelfulle opplysninger når de skal registreres i saksbehandlersystem, databaser etc. Det kan f.eks. være feil i fødselsdato, da denne ikke alltid har samme betydning i andre land og kulturer. På den måten kan personer bli registrert med feil eller mangelfulle opplysninger som gjør at tildelt informasjonsbærende personidentifikator ikke blir riktig og kanskje må endres i ettertid. Slike endringer skaper merarbeid samt fare for at data knyttet til en person kan bli vanskelig å spore eller finne fram til. Det vises også til avsnitt 7.7.4.

Mange personer kan synes det er en privat opplysning når man er født. I Norge i dag benyttes fødselsnummeret i mange sammenhenger og man blir derfor av og til indirekte tvunget til å avgi informasjon om fødselsdato og dermed alder. Enkelte personer ønsker av denne grunn at fødselsnummeret er informasjonsløst.

Det er relativt enkelt å utforme alternative personidentifikatorer som har tilstrekkelig kapasitet for en periode fram til år 2150 og lengre.

Enkelte mener at dagens fødselsnummer er for enkelt å kunne konstruere (finne fram til) da man ut i fra fødselsdato og kjønn bare har igjen individnummeret som ukjente sifre. Imidlertid er ikke fødselsnummeret ment til annet enn identifisering (påvise identiteten), og ikke til autentisering, dvs. å bekrefte eller kontrollere noen som autentisk/ekte. Autentisering kan skje ved personlige frammøte med gyldig legitimasjonspapirer eller elektronisk gjennom godkjente prosedyrer. Det er påpekt fra enkelte interessenter at en informasjonsløs identifikator kan bidra til å redusere omfanget av identitetstyverier. I dag er rutine for enkelte prosesser mangelfulle, slik at en persons fødselsnummer feilaktig kan gi tilgang til opplysninger eller rettigheter som er

uønsket. Autentiseringen er i slike tilfeller mangelfull eller fraværende. Med gode rutiner for autentisering er det liten verdi i at en informasjonsløs personidentifikator er vanskeligere å konstruere eller å huske. Fødselsnummeret er ikke hemmelig, noe som underbygger behovet for andre kilder til autentisering. I Grunndatarapporten foreslo arbeidsgruppen f.eks. at kontaktopplysninger, som bl.a. personidentifikatoren til den enkelte person, skulle være åpent tilgjengelig for enhver⁵.

Oppsummert kan argumentene og dermed behovet for en informasjonsløs personidentifikator oppsummeres slik:

- Økt stabilitet og robusthet for endringer
- Høy kapasitet
- Prinsipielt uheldig at identifikatoren inneholder informasjon
- Redusere antall dubletter. Dette tillegges liten vekt ettersom dette er et problem som best kan løses/reduseres ved andre tiltak (rutiner)
- Klargjøre identifikatorens bruksområde og egenskaper, som er identifisering og ikke autentisering.

Informasjonsbærende personidentifikator

Virksomheter innen bank- og finansnæringen, helsevesenet, skolevesenet, Forsvaret, valgmanntall, e-ID etc. benytter informasjon fra dagens fødselsnummer. Det kan være i forbindelse med saksbehandling/rettigheter der alder har betydning, vurdering av en kunde i forbindelse med lån, behandling av pasienter, opptak i skoler, militæret etc.

Dette innebærer at mange store tunge datasystemer baserer seg fullt og helt på opplysninger om fødselsdato og kjønn fra fødselsnummeret. En omlegging til informasjonsløs personidentifikator vil kunne bety svært store kostnader.

I en del sammenhenger vil informasjonsbærende personidentifikator med opplysninger om alder og kjønn være tidsbesparende da saksbehandler, behandler etc. har direkte tilgang til vitale opplysninger. Dette er tilfelle i et stort antall arbeidssituasjoner hver eneste dag, og verdien av tidsbesparingen kan være betydelig.

Svært mange husker i dag sitt fødselsnummer. De seks første sifrene, som er fødselsdato, kjenner normalt alle, og dermed gjenstår i realiteten kun behov for å huske 5 sifre. Mange opplever dette som brukervennlig.

Oppsummert kan argumentene og dermed behovet for en informasjonsbærende personidentifikator oppsummeres slik:

- Brukervennlig og lett å huske
- Redusere antall feil
- Mer kostnadseffektiv
- Gir raskere tilgang til vitale personopplysninger, som alder og kjønn
- Mindre endring fra dagens system

⁵ Dette temaet behandles grundigere i Forstudierapporten (3).

Kvalitet

Svært mange brukere er avhengig av høy kvalitet og tillit til Folkeregisteret og dermed personidentifikatoren. Den brukes i vitale arbeidsprosesser både i offentlig og privat sektor og angår alle privatpersoner i landet. Bruken av Folkeregisteret er økende, noe som gjør stadig flere avhengig av et velfungerende register.

Samfunnet blir stadig mer elektronisk og med økende informasjonsmengder. Mye informasjon som før ble formidlet muntlig, sendt på papir mellom virksomheter og fra virksomheter til privatpersoner, overføres nå elektronisk. Fødselsnummeret benyttes som identitetsnøkkel i stadig økende grad og har blitt en sentral felleskomponent.

Norge har kommet godt i gang med å få på plass elektroniske tjenester og prosesser, men det er store muligheter der det fortsatt er muntlig, delvis manuelle (diskett, tape etc.) eller papirbasert kommunikasjon. Det ligger derfor potensielt store gevinstmuligheter ved økt elektronisk informasjonsflyt som på sikt kan forventes realisert.

Denne utviklingen setter store krav til kvaliteten på opplysninger i Folkeregisteret og i særdeleshet til personidentifikatoren.

Dagens fødselsnummer har en god kontrollsifferordning hvor svært mange feilregistreringer vil bli oppdaget. Med to kontrollsifre oppdages feil som f.eks. ombytting av sifre, ikke gyldige verdier etc. Andre sammenlignbare land benytter mellom 0 til 2 kontrollsifre. F.eks. har både Sverige og Danmark ett kontrollsiffer, der Danmark også aksepterer uten kontrollsifre hvis kapasiteten blir for liten på enkeltdager (dvs. kontrollsifferet går over til å bli individnummer).

Selv om den digitale utviklingen gjør at fødselsnummeret ofte registreres automatisk, er det fremdeles behov for kontrollsifre og en god kontrollberegning. Fortsatt registreres fødselsnummeret ofte manuelt, blant annet gjennom flere selvbetjeningsløsninger.

Erfaring viser at det noen ganger oppstår feil når person går fra D-nummer til fødselsnummer, samt at enkelte har flere D-nummer. Ulike nummer lever ved siden av hverandre som dubletter. Slike dubletter skaper merkostnader for virksomhetene og problemer for de registrerte, samt gjør at utveksling av informasjon mellom virksomheter er vanskeligere og informasjonen får dårligere kvalitet. Dubletter kan også åpne opp for misbruk av f.eks. trygdeytelser. Her er det først og fremst tildelingsrutinene som er dårlige og ikke egenskapene til fødselsnummeret. Denne problemstillingen fjernes ikke alene ved en overgang til en annen personidentifikator. Inkludering av D-nummer populasjon vil imidlertid kunne redusere denne utfordringen. For D-nummer registreres det færre opplysninger samtidig som opplysningene ofte har en lavere kvalitet enn tilsvarende for fødselsnummeret. Dette skaper problemer i saksbehandlingen og for utveksling av informasjon mellom virksomheter⁶.

Det anvendes flere personidentifikatorer og nummerserier i Norge i dag i tillegg til fødselsnummeret. Vi kan nevne D-nummer, DUF-nummer, Hjelpenummer i helsesektoren, organisasjonsnummer, internasjonalt trygdenummer, telefonnummer, bankkontonummer etc. Noen av disse identifikatorene formidles mellom virksomheter og personer. En ny personidentifikator bør ha god sameksistens med eksisterende nummerserier av betydning for å unngå forveksling når dette kan skape problemer. I behovsanalysen har det kommet fram at en ny personidentifikator ikke bør kunne forveksles med hjelpenummer i helsesektoren, bankenes kundenummer, D-nummer og DUF-nummer.

⁶ Det er påpekt en del mangler for D-nummerregisteret slik dette fungerer i dag (jf. Forstudien for nytt Folkeregister og Rapport fra strategigruppen, 2011, s. 15).

Databaser

Flere fagmiljøer påpeker at det er uheldig å benytte personidentifikatoren som "primærnøkkel" i en database, fordi en person da logisk sett ikke kan bytte identifikator. Hvis identifikatoren inneholder informasjon slik som i dag, vil den være ekstra endringsutsatt. Det anbefales å benytte en egen databasegenerert primærnøkkel som f.eks. standarden GUID, globally unique identifier. Samtidig bør personidentifikatoren legges som et informasjonsfelt som er designet for oppslag, søk, etc.

De fleste store virksomheter benytter denne databasestrukturen i dag. Men noen virksomheter sitter med databaser hvor personidentifikatoren er primærnøkkel. For disse virksomhetene kan endring av personidentifikator få store konsekvenser. Sett i forhold til dagens bruk er dette systemer med behov for fornyelse, hvor overgang til ny identifikator vil kunne bidra og framskynde et slikt bytte.

For å øke robusthet for endringer kan feltet for personidentifikatoren f.eks. defineres som alfanumerisk og ha en større lengde enn 11 posisjoner (Agresso benytter f.eks. et alfanumerisk felt på 20 posisjoner).

Samhandling

Samhandling er omfattende i dag og det er et tverrpolitisk ønske om økt elektronisk samhandling. I utveksling av personopplysninger mellom virksomheter, men også innad i virksomheter er fødselsnummeret i stor grad brukt som primærnøkkel. Innføring av ny PID vil derfor være en stor risikodriver, for eksempel må det gjennomføres en omfattende koordinering mellom virksomhetene og innad i virksomhetene for bl.a test. For eksempel sendes det ca. 400 000 elektroniske meldinger i helsesektoren pr dag, og antall meldinger vil øke kraftig etter hvert som vi får et fulldigitalisert helsevesen. Det er ikke sannsynlig at all informasjonsutveksling i fremtiden vil benytte meldingsutveksling som metode, men PID vil være en primærnøkkel for informasjonsdeling om personopplysninger. At informasjonsflyten er entydig og følger riktig person, er kritisk for bl.a. pasientsikkerheten. I tillegg er informasjonsflyten mellom sektorer av vesentlig betydning både for å oppfylle innbyggernes rettigheter til tjenester fra det offentlige, men også for å nå målene om en mer effektiv offentlig forvaltning.

Overgangsordninger

Hensikten med en overgangsordning er å minimere kostnader og ulemper for virksomheter og privatpersoner. De ulike alternativene som drøftes i mulighetsstudien og alternativanalysen, vil kunne ha svært forskjellige behov for overgangsordninger.

Generelt kan sies at virksomheter som har egne store IKT-løsninger, ytrer ønske om en overgangsperiode på mellom 5 til 10 år hvis overgangen til ny identifikator innebærer systemendringer av litt størrelse. Noen få virksomheter, samt helsesektoren, har indikert at over 10 år kunne være å foretrekke. Dette er en nødvendig periode fra vedtak om en ny personidentifikator er fattet, til å kunne planlegge, utvikle og implementere ny løsning. Ved en såpass lang overgangsperiode kan kostnadene trolig også reduseres, da en del av endringene kan inkluderes i en naturlig oppgradering, utskifting etc. av systemer. For enkelte alternative personidentifikatorer som innebærer at eksisterende befolkning får nytt nummer på nytt format, er det fra flere ytre ønske om sentral oppslagsdatabase der det kan gjøres oppslag på og uttrekk både av gammel og ny personidentifikator og med kobling imellom.

Enkelte har også antydnet ønske om at eksisterende og ny personidentifikator kan leve ved siden av hverandre i en lengre overgangsperiode, i så mye som hundre år. Dermed kan man for eksisterende befolkning benytte dagens løsning og fra startdato av ny identifikator registrere

alle etter ny løsning. Her er det ulike syn hos interessentene. Enkelte, f.eks. Skatteetaten som er ansvarlig for Folkeregisteret, signaliserer økte kostnader ved en slik parallell løsning. Noen ønsker en slik parallell periode relativt kort, for å kunne fase ut gammel løsning og spare kostnader.

Leverandører av standardsystemer uttrykker at de kan gjøre de nødvendige endringer i sine systemer på under 5 år.

For virksomheter som baserer seg på standardsystemer er behovet for overgangsordninger mindre. Det forutsettes at systemleverandøren vil levere oppgraderte versjoner av systemene tilpasset ny personidentifikator og med konverteringsløsninger for eksisterende data hvis nødvendig. Virksomhetene vil trenge en tilstrekkelig tidsperiode for implementering etter at leverandører av standardsystemer har ferdigstilt nye løsninger.

Privatpersoner signaliserer behov for kun en kortere overgangsperiode. Unntaket er hvis alle skal skifte til ny personidentifikator som også innebærer skifte av legitimasjonspapirer, bank- og kredittkort, førerkort etc. I et slikt scenario kan det være behov for en lengre overgangsperiode. Dette vil også gjelde for virksomheter som f.eks. utsteder legitimasjonspapirer, kort etc.

3.7 Prosjektutløsende behov

Med det prosjektutløsende behovet menes det samfunnsbehovet som utløser gjennomføring av et tiltak til et bestemt tidspunkt.

Generelt kan prosjektutløsende behov ofte splittes inn i tre kategorier:

- Problemer som må løses
- Krav som følger av eksterne rammebetingelser
- Muligheter som kan realiseres

Det er på det rene at kapasiteten, uttrykt som tilgjengelige unike verdier per dato, er i ferd med å bli oppbrukt. Dagens fødselsnummer er bygd opp og regulert slik at det enkelte nummer er låst til et bestemt århundre. Dette for å tilfredsstille kravet om entydig identifisering og at tidligere tildelte fødselsnumre til nå døde personer ikke skal brukes om igjen.

Dagens fødselsnummer og tilgjengelig kapasitet er beskrevet i kapittel 3.1, mens behovet for nye fødselsnummer er redegjort for under kapittel 3.5.

Etter 2054 vil det bli gjenbruk av fødselsnumre idet man må ta i bruk fødselsnumre som allerede er tildelt fra 1855 og utover. Fra 2055 har dagens fødselsnummer under enhver realistisk omstendighet for liten kapasitet.

Hvis reserveserien er benyttet (individnumre 900-999), vil det etter 2039 av kapasitetshensyn trolig ikke lenger være mulig entydig å fastslå hvilket århundre en personidentifikator viser til. I Statistisk sentralbyrås prognoser ved middels nasjonal vekst vil antall fødte inkludert innvandring trolig være mindre enn tilgjengelig kapasiteten for året 2040. Selv om reserveserien skulle være ledig, vil behovet omtrent tilsvare tilgjengelig kapasitet, idet den så vidt ligger over. Ved høy nasjonal vekst vil behovet klart overstige kapasiteten i den tilgjengelige nummerserien fra 2040.

Det vurderes som nødvendig å ha noe ledig kapasitet å gå på for å kunne håndtere topper eller mindre endringer i forutsetninger etc.

På bakgrunn av drøftingen over vurderes det som sannsynlig at kapasiteten er for liten fra og med 2040.

I den videre planlegging forutsettes at ny personidentifikator derfor må være klar til bruk senest ved utgangen av 2039, dvs. tatt i bruk senest fra og med 1. januar 2040.

Det prosjektutløsende behov kan formuleres slik:

Kapasiteten til dagens fødselsnummer blir for liten fra og med 1. januar 2040 og må derfor innen dette tidspunkt være erstattet av en ny personidentifikator med tilstrekkelig kapasitet.

Hvis det i perioden fram til 2040 skulle skje utvidelse av registerpopulasjonen, dvs. endrede forutsetninger, vil tidspunktet for når dagens fødselsnummer må erstattes, kunne bli vesentlig tidligere. Det vil for noen scenarier bety fra det tidspunkt utvidelsen av registerpopulasjonen finner sted, avhengig av hvor stor utvidelsen blir. Det er behov opp mot kapasiteten gjennom tilgjengelige unike personnumre per dato som blir avgjørende for når en ny personidentifikator må være tatt i bruk.

Det prosjektutløsende behov er primært under kategorien problemer. Problemene er skapt på bakgrunn av gitte rammebetingelser som begrenset kapasitet i dagens fødselsnummer, krav til entydig identifisering av landets befolkning og en forventet befolkningsvekst.

3.8 Oppsummering av behov

Nedenfor er de viktigste behovene satt opp i prioritert rekkefølge:

1. Kapasiteten til dagens fødselsnummer er beregnet å bli for liten fra og med 1. januar 2040 og må derfor innen dette tidspunkt være erstattet av en ny personidentifikator med tilstrekkelig kapasitet. Dette er tiltakets prosjektutløsende behov.
2. Med dagens regelverk og populasjon må en ny personidentifikator ha følgende kapasitet:
 - Informasjonsbærende: Minst 900 ulike verdier per dato i 2150
 - Informasjonsløs: 37 millioner unike identifikatorverdier fram til 2150
3. Stadig økende bruk av Folkeregisteret, samt fremvekst av elektroniske tjenester og elektronisk samhandling, setter store krav til kvaliteten på opplysninger i Folkeregisteret og i særdeleshet for personidentifikatoren.

Mandatet for KVUen legger til grunn dagens lover og regler for definisjon av populasjonen.

Usikkerhet i forutsetninger, prognoser og mulighet for at spesielle hendelser inntreffer, tilsier gode kapasitetsmarginer for å dekke eventuelt økt behov.

Det er registrert interessekonflikt knyttet til informasjonsløs kontra informasjonsbærende personidentifikator. Hovedargumentene for de to typer identifikatorer er:

Informasjonsløs personidentifikator:

- Økt stabilitet og robusthet for endringer
- Høy kapasitet
- Prinsipielt uheldig at identifikatoren inneholder informasjon
- Klargjøre identifikatorens bruksområde og egenskaper, som er identifisering og ikke autentisering.

Informasjonsbærende personidentifikator:

- Brukervennlig og lett å huske

- Redusere antall feil
- Mer kostnadseffektiv
- Gir raskere tilgang til vitale personopplysninger som alder og kjønn
- Mindre endring fra dagens system

Ved innføring av ny personidentifikator er det behov for en hensiktsmessig overgangsordning som minimerer kostnader og ulemper for virksomheter og privatpersoner. Hvis overgangen til ny identifikator innebærer systemendringer av litt størrelse, bør overgangsperioden være minst 10 år.

Det er identifisert et behov for å se fødselsnummeret i sammenheng med D-nummer. Det pågående *Program for modernisering av Folkeregisteret* vurderer mulighetene for å utvide dagens populasjon til Folkeregisteret til også å inkludere D-nummer. Hvis beslutningsprosessen her resulterer i at populasjonen skal endres, vil det kunne medføre endrede krav til kapasitet for ny personidentifikator. Tidspunktet for når økt kapasitet må være tilgjengelig kan bli lik tidspunkt for utvidelse av populasjonen. Kapasitet vurderes i egen robusthetsanalyse i kapittel 7.

Ved utvikling av nye databaser anbefales benyttet databasegenererte primærnøkler, mens personidentifikatoren legges i et informasjonsfelt som er designet for oppslag, søk etc. og med robusthet for fremtidig endringer.

4 Strategi

I dette kapitlet formulerer vi samfunns mål og effektmål for tiltaket. Det gjøres rede for eventuelle målkonflikter.

4.1 Innledning

I Finansdepartementets veileder nr. 10 Målstruktur og målformulering er det gitt bl.a. følgende krav til målstruktur og målformulering:

- Det må være en klar sammenheng mellom behovet som utløser tiltaket vist i behovsanalysen og prosjektets mål. Målene må ivareta:
 - Relevans: Graden av samsvar mellom tiltakets mål på den ene siden og brukernes/samfunnets behov, interesser og prioriteringer på den andre siden
 - Konsekvens: Effekten for brukere og virkninger for samfunnet som kan tilbakeføres til tiltaket
 - Levedyktighet: I hvilken grad tiltaket fortsatt bidrar til realisering av effektmål og samfunns mål etter at prosjektets er ferdigstilt og gjennom den antatte levetiden
- Mål skal i forkant presisere hva man vil oppnå og i etterkant kunne kontrollere i hvilken grad en har greid det.
- Målene må være forståelige og målbare, dvs. såkalt SMARTE: Spesifiserte, målbare, akseptert, realistiske, tidssatte og enkle.
- Mål skal uttrykke en ønsket tilstand eller resultat, ikke aktiviteter eller arbeidsoppgaver.
- Samfunns mål skal ikke være mer generelle enn at realiseringen til en viss grad kan tilbakeføres til prosjektet. Samtidig skal det være tilstrekkelig overordnet til at det gir en god begrunnelse for tiltaket og rom for vurdering av alternative konsepter.
- Samfunns målet skal reflektere det prosjektutløsende behovet og ivareta eierperspektivet.
- Effektmålene skal være realistisk oppnåelige innen en angitt tidsramme. Effektmålene som skal ivareta brukerperspektivet, må være konsistente ved at de bygger opp under samfunns målet. Dersom effektmålene nås, bidrar dette også til at samfunns målet nås.

4.2 Samfunns målet for tiltaket

Samfunns målet skal være prosjektspesifikt, og det skal gi uttrykk for den nytte og verdiskapning tiltaket skal føre til for samfunnet. Målet skal vise eiers intensjon og ambisjon med tiltaket.

Når vi skal formulere samfunnsmålet til en ny personidentifikator, finner vi ikke dette i tidligere stortingsdokumenter eller andre formelle dokumenter da formålet til Folkeregisteret ikke er entydig definert og forankret. I det pågående strategiarbeidet konkluderes det med at et nytt lovverk bør inkludere en tydelig formålsparagraf (8). Folkeregisterets rolle og funksjoner er imidlertid drøftet og beskrevet i flere av referansedokumentene.

Personidentifikatoren er den primære nøkkelen i Folkeregisteret, og benyttes for de fleste oppslag i Folkeregisteret og ikke minst ved utveksling av informasjon mellom virksomheter.

Grunddatarapporten oppsummerer Folkeregisterets formål på følgende måte:

Innsamling, lagring og distribusjon av grunndata om personer med tilknytning til Norge for å:

- Følge opp enkeltpersoners rettigheter og plikter
- Ivareta samfunnsfunksjonene
- Bidra med informasjon til samfunnet
- Skape oversiktighet om bruken av personinformasjon

Tildeling av entydig identifikasjon til personer registrert i registeret

Bidra til sikker og kostnadseffektiv samhandling i og med offentlig sektor

I strategirapporten til *Program for modernisering av Folkeregisteret* (8) er målene til et framtidig Folkeregister omtalt som følger:

“Et fremtidig personregister bør være et enhetlig register, en ”master” for persondata som dekker relevante og viktige samfunnsbehov og utgjøre et grunddataregister for personer på samme måte som Enhetsregisteret er for juridiske personer med organisasjonsnummer.

Generelt skal et register være heldekkende, kvalitetssikret, oppdatert, dokumentert og brukertilpasset. Et personregister skal inneholde data av god kvalitet og aktualitet og med god dekningsgrad for alle deler av registerpopulasjonen (personkretsene).

.....

Det er et overordnet mål at det finnes ett personregister i Norge for alle personer som norske myndigheter har behov for å identifisere. Dagens ordning med ett folkeregister over bosatte, en D-nummerpopulasjon uten sentral oppfølging og en egen utlendingsdatabase gir en rekke ulemper.”

Med bakgrunn i mandatet, Grunddatarapporten og Rapport fra strategigruppen til modernisering av Folkeregisteret formuleres følgende samfunnsmål for ny personidentifikator i Folkeregisteret:

En unik personidentifikator egnet for praktisk anvendelse til minst år 2150.

4.2.1 Effektmål

Tabellen nedenfor angir effektmål for en ny personidentifikator. For det enkelte mål er det knyttet en eller flere måleindikatorer for å gjøre effektmålene etterprøvbare (jfr SMARTE-prinsippet):

	Effektmål	Indikator
E1	Tilstrekkelig kapasitet for dagens registerpopulasjon og robusthet for eventuell framtidig utvidelse (D-nummer).	Tilstrekkelig kapasitet, også for en utvidelse av populasjonen. For informasjonsbærende identifikatorer: Antall tilgjengelige løpenummersverdier per dato. For informasjonsløse identifikatorer: Totalt antall unike verdier minst fram til år 2150.

E2	Kostnadseffektiv realisering og drift	Estimerte kostnader (både prissatt og ikke prissatt).
E3	Ha en utforming som oppleves brukervennlig.	Er lett å huske (ikke prissatt). Tidsbruk ved registrering/tasting (ikke prissatt) Tidsbruk ved identifisering og fremskaffelse av personopplysninger som alder og kjønn (ikke prissatt). Det er enkelt å rette feil/endre en persons identifikator (ikke prissatt).

Tabell 4: Effektmål og måleindikatorer

Tilstrekkelig kapasitet for dagens registerpopulasjon og krav om unikhet fram til 2150 som angir det prosjektutløsende behovet, er inkludert som absolutt krav i kravkappitelet.

E1 om kapasiteten til registerpopulasjonen vurderes i den samfunnsøkonomiske analysen gjennom en robusthetsanalyse hvor man ser på følgende:

- Dagens populasjon
- Dagens populasjon og D-nummer

E2 er i all hovedsak prissatt og inkludert i den samfunnsøkonomiske analysen. Noen kostnadselementer er ikke prissatt og er dermed behandlet i som ikke prissatt effekt i analysen.

E3 treffer både virksomheter og privatpersoner. Alternativene skiller seg fra hverandre og vil bli rangert på indikatorene med score som ikke-prissatte virkninger.

Behovet for høy kvalitet på opplysninger i Folkeregisteret og i særdeleshet på personidentifikatoren slik at bl.a. feil minimaliseres, forutsettes ivaretatt gjennom effektmål 3.

Personvern, sikkerhet og kontrollrutiner er tatt inn som krav i kravkappitelet.

4.2.2 Målkonflikter

Vi har identifisert følgende målkonflikt:

Det er en potensiell målkonflikt mellom en brukervennlig identifikator som er lett å huske og en identifikator som har tilstrekkelig stor kapasitet. En brukervennlig identifikator kan bygge på informasjonsbærende elementer, slik dagens fødselsnummer som består av 6 tall som markerer fødselsdato. Dette er enkelt å huske, men legger begrensning på antall tallkombinasjoner.

5 Krav

Her i kravkapittelet sammenfattes de betingelsene som skal oppfylles ved gjennomføringen av tiltaket. Kravene er begrenset til relevante krav for tiltaket og for en avgrensning av mulighetsrommet.

5.1 Innledning

I Finansdepartementets rammeverk er det blant annet stilt krav til at, sitat:

“Det overordnede kravkapitlet skal sammenfatte betingelsene som skal oppfylles ved gjennomføringen. Det er tale om to typer krav:

- *Krav som utledes av samfunns- og effektmålene.*
- *Ikke-prosjektspesifikke samfunns mål. I praksis vil slike mål fremstå som rammebetingelser for tiltaket. Av denne grunn er det mest hensiktsmessig å behandle disse målene i kravkapitlet. Da det finnes svært mange generaliserte mål, må antallet som analyseres begrenses til slike som er spesielt relevante for undersøkelsen av mulighetsrommet.*

Kravkapitlet skal være fokusert mot effekter og funksjoner. I forhold til det å ha en konsistent prioritering og robusthet i dataenes utsagnskraft på et overordnet nivå, er teknisk løsningsorientering og detaljeringsgrad av underordnet betydning.”

Kravene skal være tilfredsstillende utledet av samfunns- og effektmålene, eller av de beskrevne rammebetingelsene for tiltaket. Videre skal det være indre konsistens mellom kravene.

Absolutte krav begrenser mulighetsrommet for hva som er relevante alternativer, dvs. at alternativer kan bli “silt bort”. Det er derfor viktig at det under absolutte krav kun tas med de helt nødvendige kravene, at disse er nøkternt formulert og forankring beskrevet.

Kravene er gruppert i følgende to grupper:

- Absolutte krav: Krav som må oppfylles
- Viktige krav: Krav som er prioritert høyt

Finansdepartementets Veileder nr. 9 Utarbeidelse av KVU/KL dokumenter (9) angir følgende sammenheng mellom krav og den samfunnsøkonomiske analysen:

“Krav som er relatert til effekter som kan omregnes til kroner inkluderes som prissatte konsekvenser i den samfunnsøkonomiske analysen. Krav som helt eller delvis ikke kan innarbeides som prissatte konsekvenser, skal behandles som tiltaksspesifikke ikke-prissatte konsekvenser i den samfunnsøkonomiske analysen.”

5.1.1 Absolutt krav:

Det stilles tre absolutte krav til ny personidentifikator, innenfor kapasitet og personvern:

	Absolutte krav	identifikator
KA1	Ha kapasitet for dagens registerpopulasjon minst fram til 2150.	Informasjonsløs: Minst 37 millioner identifikatorverdier.

	Absolutte krav	identifikator
		Informasjonsbærende: Minst 900 ulike verdier per dato i 2150.
KA2	Hver person skal ha en unik personidentifikator.	Det skal ikke forekomme at flere personer får tildelt samme identifikatorverdi.
KA3	Oppfylle gjeldende regelverk for personvern og sikkerhet.	Ingen brudd på gjeldende regelverk.

Tabell 5: Absolutte krav

KA1 Følger av kapittel 3.5 Etterspørselsbaserte behov, KA2 og KA3 følger av kapittel 3.4 Normative behov.

Absolutt krav til kapasitet legger til grunn dagens lover og regler for definisjon av populasjonen slik det er definert i mandatet. For informasjonsløs personidentifikator uttrykkes kravet gjennom det kumulative behovet, og for informasjonsbærende i tilgjengelige løpenummer per dato. Eventuelle endringer i populasjonen og dermed endrede krav til kapasitet, er behandlet gjennom en robusthetsanalyse i den samfunnsøkonomiske analysen.

5.1.2 Viktige krav

Det er inkludert 4 viktige krav som er vurdert sentrale for at vi skal få et velfungerende Folkeregister med ny personidentifikator:

	Viktige krav	Identifikator
KV1	Lav gjennomføringsrisiko	Grad av kompleksitet og gjennomførbarhet for virksomhetene i prosjektperioden. En hensiktsmessig overgangsordning som minimerer kostnader og risiko.
KV2	Høy brukervennlighet	Om personidentifikatoren er lett å huske: - Nyregistrering - Eventuell overgang fra dagens fødselsnummer. I hvilken grad privatpersoner må endre identifikasjonspapirer eller foreta andre endringer knyttet til personidentifikator. Alle skal ha en unik og riktig personidentifikator.
KV3	Gode kontrollrutiner	Bruk av kontrollsiffer Sannsynlighet for feil.
KV4	God sameksistens med andre nummerserier	Identifiserte konflikter med andre nummerserier.

Tabell 6: Viktige krav

6 Mulighetsstudie

For å få en best mulig vurdering av mulighetsrommet for større statlige investeringer, stilles det krav i KVUer om gjennomføring av en mulighetsstudie. Denne skal være bred, i den forstand at fremgangsmåtene og alternativene som behandles også skal dekke mindre opplagte og mer utradisjonelle fremgangsmåter og tankesett. Mulighetene skal følge av behov/mål/krav og det skal dermed være indre konsistens.

Det er her tatt utgangspunkt i effektmålene som er formulert i kapittel 4.2.1. Dette innebærer at mer radikale måter å tenke på som andre måter å organisere samfunnet på, ikke er aktuelle. Det er heller ikke vurdert som aktuelt at fødselsnummer ikke lenger skal være unikt. Dette ville i så fall ha redusert eller eliminert oppnåelsen av hovedoppgaven å fungere som en personidentifikator med unike verdier.

Som et ledd i mulighetsstudien har det blitt gjennomført to idedugnader i oktober 2011. Her ble det foruten personer fra prosjektet også inkludert eksterne interessenter fra NAV, SSB, Brønnøysundregistrene og Finansnæringens fellesorganisasjon. I tillegg var det i forkant av dette møter med andre viktige interessenter, deriblant Nordea, Toll- og avgiftsdirektoratet, DNB, Utlendingsdirektoratet, Sparebank1, NAV og Helsedirektoratet. I tillegg har temaet vært diskutert i referansegruppen til prosjektet.

6.1 Gjenbruk av fødselsnumre fra døde personer

En mulighet for å øke kapasitetene ville vært gjenbruk av gamle fødselsnumre fra døde personer. Denne anser vi imidlertid som ikke å gi unike numre og bryter dermed med forutsetningene i Folkeregisterloven (10).

6.2 Mulighetsdimensjoner

Det har blitt gjort en bred vurdering knyttet til mulighetsdimensjoner, da spesielt andre måter enn en numerisk serie som kan brukes som unik identifikator. De begrensningene som er satt er disse:

- Fødselsnummeret må være unikt.
- Fødselsnummeret må kunne kommuniseres synkront og asynkront.
 - Synkront: Øyeblikkelig tilbakemelding
 - Asynkront: For eksempel papirskjema

6.2.1 Alfnumerisk

Det å bruke alfanumerisk identifikator øker kapasiteten innenfor en gitt lengde (antall posisjoner) betraktelig. Dette kan bestå i at en eller flere posisjoner i tillegg til et tall også kan være en bokstav og/eller et tegn. Den faktiske økningen i kapasitet avhenger av hvor mange bokstaver og tegn man tillater å bruke, samt hvor mange av posisjonene som får denne egenskapen.

6.2.2 Andre egenskaper enn numre/bokstaver

En variant av numerisk eller alfanumerisk identifikator er å bruke andre egenskaper i tillegg til selve numrene/bokstavene. Eksempler på dette er

- Hvor hardt man trykker
- Hvor lenge man holder knappen inne
- Hvor fort man trykker

Morsekode er en gammel metode som utnytter denne dimensjonen.

6.2.3 Biometri

En mulig dimensjon er bruk av biometri⁷, da spesielt fingeravtrykk, ansiktsgjenkjenning eller øyeiris. Det går også an å bruke kombinasjoner av numerisk eller alfanumerisk identifikator og biometri.

6.2.4 Informasjonsløs identifikator

Dagens identifikator inneholder opplysninger om fødselsdato (første 6 sifre) og kjønn i 9. posisjon (partall betyr hunkjønn og oddetall hankjønn). En mulighetsdimensjon er å ikke videreføre dette, dvs. gå over til en informasjonsløs identifikator. For øvrig vil det at man i det hele tatt har fødselsnummer gi informasjon om tilknytning til Norge.

Det er ikke identifisert andre opplysninger som bør ligge i fødselsnummeret som gjør at fødselsnummeret bør bli "mer informasjonsbærende".

6.2.5 Antall kontrollsifre

I dagens løsning er det lagt inn 2 kontrollsifre (10. og 11. siffer). Dette er flere kontrollsifre enn mange tilsvarende systemer bruker og det kan derfor være aktuelt å endre kontrollberegningen eller redusere antallet. Eventuelt kan man endre kontrollen på andre måter som gir økt kapasitet.

6.2.6 Sektorvis identifikator

Sektorvis identifikatorer innebærer at personer ikke har en felles personidentifikator, men flere ulike, avhengig av tjenesteområde. En person har f.eks. en identifikator for helsesektoren, en for politiet, en for kommunale tjenester osv. I Europa praktiseres denne løsningen av Østerrike.

6.2.7 Konvertering

Dersom man velger en løsning som avviker betydelig fra dagens, kan det være aktuelt å konvertere dagens fødselsnumre til nytt format, det vil si at hele populasjonen skifter personidentifikator.

6.2.8 Lengde

En personidentifikator kan ha ulike lengder, her er det snakk om antall posisjoner.

⁷ Wikipedia: "Biometri er måling av biologiske mønstre. Biologiske mønstre kan være fysiologiske karaktertrekk (for eksempel fingeravtrykk) eller adferdsmønstre (for eksempel ganglag)."

6.3 Vurdering av mulighetsdimensjoner

6.3.1 *Alfanumerisk*

Bokstaver har i stor grad de samme egenskapene som tall i denne sammenhengen og kan være en hensiktsmessig måte å øke kapasiteten på (6). Med utgangspunkt i det norske alfabetet er det noen begrensninger. Bokstaven I er til forveksling lik tallet 1. Bokstaven O er lik tallet 0 og Æ, Ø og Å kan minne om E, O og A, samt at bokstavene ikke er å finne på svært mange tastaturer. Andre lands alfabeter (gresk, arabisk, kinesisk, et cetera) er lite aktuelle da de fleste vil ha problemer med å kunne lese opp og/eller skrive disse. Symboler (for eksempel &, \$, *, et cetera) kan vurderes men gir komplikasjoner sammenlignet med bokstaver da de er mindre tilgjengelige, mindre utbredt og kan skape forvirring i bruk.

Bruk av et sett bokstaver i tillegg til tall sees på som realistisk og behandles videre i alternativanalysen.

6.3.2 *Andre egenskaper enn numre/bokstaver*

Egenskaper knyttet til tid og styrke vil ikke kunne kommuniseres på en lett og entydig måte asynkront (papirskjema). Det er dessuten begrensninger knyttet til valg av teknologi, for eksempel vil de fleste av dagens tastaturer ikke kunne brukes for å registrere styrke på et tastetrykk.

Det vurderes ikke som realistisk å innføre dette og denne dimensjonen tas derfor ikke inn i alternativanalysen.

6.3.3 *Biometri*

Biometri er i noen grad i bruk som identifikator i samfunnet, da spesielt knyttet til autentisering. På en del PCEr kan man bruke fingeravtrykk for å logge på, i noen få biler kan man starte motoren på denne måten. Det er også andre eksempler på autentisering ved hjelp av fingeravtrykk. Noen mobiltelefoner tillater bruk av ansiktsgjenkjenning som alternativ til kode. Øyeskanning er også i noen grad tatt i bruk.

Det er imidlertid betydelige utfordringer i å bruke biometri som grunnlag for en entydig personidentifikator.

En stor fordel og samtidig et problem med en biometrisk basert personidentifikator, er at den vil identifisere en person entydig på verdensbasis. Dette reduserer muligheten for å tildele en person ny identitet, noe en ikke kan utelukke som et legitimt behov.

Det største problemet ligger i at biometri er svært vanskelig å bruke asynkront som for eksempel på et papirskjema. Selv i tilfeller der dette lar seg registrere (som for eksempel ved å legge på et fingeravtrykk) så kan det ikke leses direkte uten maskinell hjelp. Også for synkron kommunikasjon er det utfordringer knyttet til overføring av data, ved et oppslag mot en sentral database vil informasjonen uansett være i digitalisert form. Det er også utfordringer knyttet til kontroll da det vil være usikkerhet om ett sett med data som ikke godkjennes, er ugyldig (forfalsket) eller bare av dårlig kvalitet. Biometriske data vil i praksis måtte oversettes til en annen identifikator som kan brukes parallelt med biometri.

I sammenheng med personidentifikator vurderes biometri som umodent. Biometri kan nok ha en del positive egenskaper, men neppe som *alternativ* til numerisk eller alfanumerisk identifikator. Se for øvrig (3) for andre vurderinger rundt biometri.

Samlet sett vurderes ikke biometri å være aktuelt nok til å bli videreført til alternativanalysen.

6.3.4 Informasjonsløs identifikator

Å fjerne informasjon om fødselsdato og kjønn vil gi noen ulemper da dette er opplysninger som i større og mindre grad er tatt i bruk av virksomheter som bruker fødselsnumre. Ett eksempel er at banker kan bruke fødselsnummer til å se at en person ikke kan ta opp lån fordi han/hun er yngre enn 18. En annen ulempe er at en informasjonsløs identifikator vil bli vanskeligere å huske for mange eiere.

Det kan også være fordeler ved å ta i bruk informasjonsløs identifikator da det kan argumenteres for et bedre personvern og at nummeret blir mer bestandig. For noen virksomheter er det usikkerhet knyttet til vitale opplysninger, som for eksempel med registrering av innvandrere der fødselsdato er ukjent.

Samlet sett anses det som aktuelt å vurdere alternativer med informasjonsløse identifikatorer og dette går inn i alternativanalysen.

6.3.5 Antall kontrollsifre

Rapporten til Selmersenteret (vedlegg 1) behandler dette temaet og foreslår noen løsninger. Det vurderes som aktuelt å endre kontrollberegningen eller redusere antall kontrollsifre for å oppnå økt kapasitet, noe som er videreført til alternativanalysen.

6.3.6 Sektorvis identifikator

Dette er en alternativ måte å organisere personinformasjon på, som noen vurderer bedre med hensyn til personvern. Det blir vanskeligere å sammenstille informasjon fra ulike virksomheter/sektorer og følgelig sette sammen flere opplysninger om enkeltpersoner.

Denne modellen avvises med bakgrunn i innspill fra viktige primærinteressenter (referansegruppen). Det er store fordeler knyttet til den norske og skandinaviske modellen med å ha en primærnøkkel for kommunikasjon og samhandling mellom virksomheter på tvers av sektorer. Et felles fødselsnummer bidrar til effektivisering av offentlig forvaltning og øker mulighetene for å utvikle digitale tjenester bl.a. annet basert på offentlige grunndata.

6.3.7 Konvertering

Dette er aktuelt i en situasjon hvor man endrer personidentifikatoren betraktelig og dette videreføres dermed til alternativanalysen.

6.3.8 Lengde

Lengde er en aktuell mulighetsdimensjon som tas videre i alternativanalysen.

6.4 Oppsummering

Følgende mulighetsdimensjoner vurderes som realistiske og videreføres i alternativanalysen:

- Alfanymerisk
- informasjonsløs identifikator
- antall kontrollsifre
- med og uten konvertering
- lengde - antall posisjoner.

Disse vil settes opp i flere alternativer og kombinasjoner.

I mandatet stilles det krav om å ha med de 2 alternativene fra grunndatarapporten, jfr. kap 2.4

"I rapporten "Utveksling av grunndata på personinformasjonsområdet" er det foreslått to alternativer for ny identifikator. Primært foreslås en såkalt informasjonsløs identifikator, mens det sekundært foreslås at dagens fødselsnummer videreføres, men at dets oppbygning endres ved å endre et av sifrene fra å være kontrollsiffer til å inngå i løpenummeret for å øke kapasiteten. Disse to alternativene skal tas med i mulighetsstudien og alternativanalysen."

Dette er tilfredsstillende ved at både den informasjonsløse identifikatoren behandles videre og at man vurderer reduksjon av kontrollsiffer eller endret kontrollberegning.

Totalt er over 40 ulike alternativer vurdert i mulighetsstudien. Gjennom en prosess med flere workshops og bred involvering, ble antallet deretter redusert til 8 alternativer.

11 sentrale interessenter og brukere av dagens fødselsnummer både fra offentlig og privat sektor, var med i en pilotstudie av disse åtte alternativene. Seltersenteret leverte samtidig en rapport hvor disse alternativene ble analysert og vurdert. Referansegruppen fikk en gjennomgang og drøftet resultatet fra nevnte gjennomganger og ga sine vurderinger og innspill. På dette grunnlag ble så antall alternativer redusert fra åtte til fire.

De fire alternativene ble så gjenstand for en ny og bred runde med konsekvenskartlegging. Det er dette som er behandlet videre og utdypet i alternativanalysen.

7 Alternativanalyse

Med bakgrunn i de foregående kapitler og i særdeleshet det identifiserte mulighetsrommet, er det i tillegg til dagens fødselsnummer (nullalternativet) identifisert fire alternative konsepter for en ny personidentifikator til Folkeregisteret, hvorav ett alternativ har to undervarianter. Alternativene blir i alternativanalysen beskrevet og bearbeidet i en samfunnsøkonomisk analyse i henhold til Finansdepartementets veileder i samfunnsøkonomiske analyser, hvor både prissatte og ikke-prissatte virkninger er tatt med. Basert på analysen gjøres det en anbefaling ut i fra hvilket alternativ som oppfyller effektmålene ut fra beregnet beste samfunnsøkonomiske effektivitet.

7.1 Beskrivelse av alternativene

Beskrivelsene av alternativene omfatter:

- Oppbygging med vekt på type og antall tegn, og om personidentifikatoren er informasjonsbærende eller ei.
- Kapasitet. For alternativene 0,1 og 2 som inneholder informasjon om fødselsdato, angis kapasitet som unike verdier per dato. For de informasjonsløse identifikatorene angis kapasitet som antall unike og gyldige verdier.
- Kontrollberegning og antall kontrollsifre.

7.1.1 Nullalternativet – dagens fødselsnummer

Oppbygging

Nullalternativet er en videreføring av dagens fødselsnummer, som har følgende oppbygging:

$D_1D_2M_1M_2I_1I_2I_3K_1K_2$

D: dato, M: måned, Å: år, I: individsiffer/personnummer, K: kontrollsifre. I_3 viser kvinne (partall) eller mann (oddetall).

Det gis informasjon om kjønn og fødselsdato, og hvor dette legger begrensninger for antall nye unike nummer for hver enkelt dag.

Kapasitet

I perioden etter 2054 opphører dagens fødselsnummer som unik identifikator fordi nummerserien da vil få overlapp med brukte nummer for personer født i perioden fra 1855. Muligheten for å identifisere hvilket århundre en person er født forsvinner fra dette tidspunktet.

Dersom dagens fødselsnummer skal eksistere etter 2054, vil det innebære at man har gjenbruk av gamle nummerserier for personer født fra 1855, og at nummeret derfor ikke er unikt og entydig. I teorien kunne en tenke seg å gjenbruke gamle nummer til personer som har vært døde i over 50 år og er født før 1900. Eventuelt å bruke gyldige, ubrukte numre fra brukte nummerserier. Likevel er ikke dette en løsning fordi kapasiteten blir uansett kritisk lav på sikt.

I behovsanalysen kapittel 3.1 har vi gjort rede for de kapasitetsutfordringer dagens fødselsnummer har som følge av befolkningsvekst. Kapasiteten framstår som kritisk lav før 2054, avhengig av hvor stor befolkningsveksten blir. I tabellen nedenfor er kapasiteten illustrert. Vi ser at etter 2054 er kapasiteten i alle tilfeller for lav.

Tidsperiode	Kapasitet per dag
2000-2039	496
2040-2054 (eks. reserve)	331
2040-2054 (inkl. reserve)	413
2054-2150	For liten

Tabell 7: Oversikt over total kapasitet for begge kjønn pr. dato

Kontrollmekanisme

0-alternativet har en meget sterk kontrollrutine gjennom bruk av to kontrollsisfre. Det benyttes modulus 11 kontroll, lengden på fødselsnummeret er på 11 posisjoner og alfabetet har 10 symboler (sifrene 0...9). Da 0 ikke kan benyttes som vekt, blir det igjen 10 unike vekter (1...10) ved beregning av kontrollsisfre. For ytterligere omtale se avsnitt 3.1, bruk av kontrollsisfre.

7.1.2 Alternativ 1 - Dagens fødselsnummer med endret kontrollsisfer

Oppbygging

Alternativ 1 er identisk med dagens fødselsnummer bortsett fra at det ene kontrollsisferet delvis frigjøres og går inn i individnummeret. Personidentifikatoren får da følgende oppbygging:

$$D_1 D_2 M_1 M_2 \ddot{A}_1 \ddot{A}_1 I_2 I_3 K_1 K_2$$

D: dato, M: måned, Å: år, I: individsiffer/personnummer, K: kontrollsisfer. I_3 viser kvinne (partall) eller mann (oddetall).

Kjønn og fødselsdato identifiseres på samme sett som tidligere. Fødselsnummer tildelt etter dagens ordning kan beholdes.

Det første kontrollsisferet K_1 vil få andre egenskaper enn i dagens fødselsnummer. Dette er nærmere omtalt nedenfor.

Kapasitet

Kapasitet er i utgangspunktet 826 per dag. Alternativet skilles fra dagens nummerserie ved å benytte S modulo 11 kontroll, der S er kontrollsummen som bestemmer de to kontrollsisfrene. I dagens fødselsnummer er K bestemt slik at S er lik 0 modulo 11. I det nye systemet kan S for første kontrollsisfer K_1 (S_1) anta verdiene fra 1 til 10. Hver verdi av S_1 fra 1 til 10 har en kapasitet på 826 per dato. På den måten kan kapasiteten på 826 mangedobles, og oppfyller kravet om kapasitet på minimum 900 pr. dag med god margin.

Totalt er kapasiteten 8260 ubrukte identifikatorer for hver DDMMÅÅ for en tidsperiode på 100 år. Det må vurderes hva som er en fornuftig fordeling av denne kapasitet over tid, dvs. hvordan man fordeler parameterverdier over tid/mellom århundrer. Det første århundret kan man for eksempel anvende S_1 lik 1 med kapasitet 826 pr. dato, alternativt S_1 lik 1 eller 2 med kapasitet 1652 pr. dato.

Kontrollmekanisme

For dagens fødselsnummer har K_1 en kontrollsum S_1 som alltid er lik 0 modulo 11. En systematisk måte å unngå "gamle" verdier er derfor i fremtiden å definere K_1 slik at den tilhørende kontrollsummen S_1 er ulik 0 modulo 11. Da får man et klart skille mellom tidligere brukte nummer og nye nummer.

I likhet med nullalternativet vil man etter 2054 i utgangspunktet ikke kunne identifisere hvilket århundre personer er født direkte fra fødselsnummeret. For å kunne kode for århundre på en systematisk måte, samt å beholde en viss feildeteksjonsevne, foreslås det å definere ulike verdier for S_1 til århundre.

Det anbefales at alternativet kombineres med nytt felt i Folkeregisteret hvor århundre for fødselsdato angis.

Alternativ 1 benytter samme feilkontroll som i dag for det andre kontrollsifferet (K_2). Dersom man på et fremtidig tidspunkt skulle ønske å utvide kapasiteten ytterligere kan man vurdere et tilsvarende grep for K_2 som for K_1 her.

7.1.3 Alternativ 2 – Alfamerisk og informasjonsbærende

Oppbygging

Alternativ to er ulikt dagens fødselsnummer ved at de tre individsifrene både kan være tall og bokstaver (alfanumeriske tegn). Personidentifikatoren har følgende oppbygging:

$$D_1 D_2 M_1 M_2 \text{Å}_1 \text{Å}_2 A_1 A_2 A_3 K_1 K_2$$

D: dato, M: måned, Å: år, A: individsiffer eller -bokstav, K: kontrollsiffer.

Det innføres alfanumeriske tegn for posisjonene $A_1 A_2 A_3 K_1 K_2$. Det forutsettes at av de 5 posisjonene for individnummer og kontrollsiffer, er minst tre posisjoner bokstaver. Kravet er satt for enklere å skille ny personidentifikator fra dagens fødselsnummer samt oppdage feilskrivning etc.

Også for dette alternativet er det mulig å skille mellom århundre og mellom kvinne og mann ved å kode de alfanumeriske tegnene. Koding av århundre og kjønn kan gjøres på flere måter, f.eks. gjennom første bokstav.

De som har fått tildelt fødselsnummer etter dagens ordning beholder dette. Kun nye borgere får det nye alfanumeriske nummeret. De to nummerseriene kan leve parallelt og vil måtte gjøre dette i lang tid (utover analyseperioden). Med følgende rutine kan man gå fram for å skille nye nummer fra dagens fødselsnummer, og bestemme hvilken kontrollmekanisme som skal anvendes.

- Det sjekkes først om det er gammel eller ny identifikator (Selmer foreslår som kriterium om det er minst tre bokstaver blant de siste fem eller ikke).
- Hvis det er identifikator fra dagens fødselsnummer, benyttes eksisterende rutine for feilkontroll. I motsatt fall benyttes ny rutine for feilkontroll.

Antall tegn i alfabetet,

Ved å utvide symbolene til også å omfatte bokstaver for de tre individposisjonene og de to kontrollsifrene, økes kapasiteten vesentlig. Videre er det en sammenheng mellom antall godkjente tegn og kontrollmekanisme.

Det vil være mange ulike kodekonstruksjoner som påvirker hvor stor kapasiteten blir. En faktor er hvor mange bokstaver i alfabetet man tillater å bruke. Det anbefales ikke å anvende

bokstaver som lett kan blandes med tall (eks. I, O, B, D etc.). Jo flere tegn man tillater å bruke jo større er sannsynligheten for symbolkonflikter.

Feildeteksjonssystemet blir konstruert med bakgrunn i antall symboler som tillates og deres posisjon. Ved å ha to kontrollsymbol og tillate maksimalt to desimale (tall) symbol i de siste fem symbolene (dvs. minst tre bokstavposisjoner) er det mulig å ha en god kontrollmekanisme som opererer i sameksistens med den gamle Selmerkoden (modulu 11).

Tegn i alfabetet
q : 17 – 17 tegn anvendes
1 5 9 A B C D E G K M P T U W X Z
q : 19 – 19 tegn anvendes
2 5 8 9 A C D E F H J L M N P T U W X
q : 23 – 23 tegn anvendes
2 5 8 9 A C D E F G H I J K L M N P R T U W X
q : 29 – 29 tegn anvendes
3 6 7 9 A B C D E F H I J K L M N O P Q R S T U V W X Y Z

Tabell 8: Beste alfabet dersom alle desimalsifre skal være med

I tabellen over har vi gitt alfabetene for $q = 17, 19, 23$ og 29 , som er aktuelle primtallsverdier. Det er hensiktsmessig å velge alfabetstørrelse som er identisk med en primtallverdi, av hensyn til moduluskontrollen. Det er anbefalt å ta i bruk 23 tegn, $q=23$ ut i fra funksjonalitet (kapasitet i forhold til symbolkonflikter).

Når det gjelder hvilke bokstaver i alfabetet som bør velges er oppstillingen over av foreløpig karakter. Valg av alfabet kan ha betydelig innflytelse på sannsynlighet for feil enkeltsymbolgjenkjenning. Dersom man velger alternativ med alfanumeriske tegn, vil det være behov for å utredes mer grundig før endelig valg av tegn foretas.

Kontrollmekanisme og kapasitet

Om antall tegn økes utover 11 må det etableres en ny modulu kontroll, for eksempel modulu 13, modulu 17, modulu 19 etc. Modulu kontrollen bygges alltid opp om et primtall, og man kan ikke velge antall tegn over dette primtallet. Sammenhengen mellom antall tegn og moduluskontroll er illustrert i den øverste tittelraden i tabellen nedenfor.

Hvordan antall symboler, moduluskontroll og maksimalt antall desimale symbol påvirker kapasiteten, er illustrert i tabellen. Se også vedlegg 2 for mer detaljert informasjon.

Max antall desimale symbol	Modulu 17 q=17, bruk av 7 bokstaver og 10 tall.	Modulu 19 q=19, bruk av 9 bokstaver og 10 tall	Modulu 23 q=23, bruk av 13 bokstaver og 10 tall
0	55	158	693
1	464	1 059	3 387

2	1 645	3 087	7 535
3	3 342	5 322	10 739
4	4 561	6 575	11 973
5	4 913	6 859	12 167

Tabell 9: Illustrasjon av hvordan symbolbruk påvirker personidentifikatorens kapasitet målt i antall pr. dag. Det er forutsatt bruk av to kontrollsifre. Q er antall tillatte symboler, som både kan være bokstaver og tall.

Den valgte løsningen, maksimalt to desimale symbol og 23 symbol totalt, har en kapasitet på minst 7535 identitetsnumre per dato (DDMMÅÅ) (se vedlegg 2).

7.1.4 Alternativ 3 – Informasjonsløst kundennummer

Hovedalternativet 3a innebærer at tildelte fødselsnumre etter dagens ordning beholdes, dvs. at personer som har fødselsnummer fram til ny identifikator trår i kraft, beholder sitt gamle.

Oppbygging

Vi har kalt dette alternativet informasjonsløst kundennummer fordi det er en tilfeldig tallrekke uten informasjon om fødselsdato eller kjønn. Det betyr også at man ikke tildeler kontinuerlig, f.eks. i en stigende eller synkende nummerserie, da det vil kunne gi opplysninger om personens fødselsdato, tidspunkt for registrering etc. Alternativet har følgende oppbygging:

I₁I₂I₃I₄I₅I₆I₇I₈I₉K₁K₂

I: individsiffer/personnummer, K: kontrollsifre.

Kapasitet

Dette alternativet har i utgangspunktet en kapasitet på 826 millioner identifikatorverdier.

Når nummerserien er tilfeldig kan det oppstå konflikt med andre typer nummerserier som også består av 11 siffer. Dette gjelder for følgende nummerserier (se også vedlegg 3):

- Nummerserier brukt som fødselsnummer etter dagens system.
- Nummer som kan forveksles med fødselsdato (første siffer 01 – 31).
- Nummerserier brukt som D-nummer (første siffer fra 41 – 71).
- Helsesektoren: Hjelpenummer
- Konstruerte kundennummer til bankene.

Ved å vaske bort disse nummerseriene reduseres kapasiteten fra 836 millioner til 123 millioner. Dette er fortsatt mer enn tilstrekkelig god kapasitet.

Folkeregisteret kan legge inn sperre for disse nummerserier slik at man unngår konflikter. Det er viktig at en slik sperre er 100 prosent korrekt.

Kontrollmekanisme

Alternativet har to kontrollnumre. Det kan benyttes samme feilkontroll (modulo 11 kontroll) som ved dagens fødselsnummer, slik at feildeteksjonsegenskapene er tilnærmet uendrede sammenliknet med 0-alternativet.

Ny identifikator for hele befolkningen

Dette alternativet har en variant 3b:

3b: Det fattes et "prinsippvedtak" om at alle skal over på informasjonsløs personidentifikator, noe som innebærer at også de som har fødselsnummer etter dagens ordning skal få ny informasjonsløs identifikator.

Når hele befolkningen skal få ny identifikator, er det behov for å benytte tekniske løsninger som forenkler overgangen. For å illustrere dette tas utgangspunkt i løsning som Folkeregisteret i dag har for personer hvor flere identifikatorer er tatt i bruk. Dette kan gjelde utlendinger som i første omgang fikk D-nummer, men som senere har bosatt seg og fått fødselsnummer. Tilsvarende har enkelte personer endret fødselsnummer grunnet feilregistrering eller ny dokumentasjon av fødselsdato. Disse personene har i dag fått tildelt to identifikatorer, en som er utgått og en ny.

Folkeregistermyndigheten tilbyr i dag tjenester som returnerer/viser nytt fødselsnummer ved søk på utgått nummer. Det finnes også abonnementstjenester for ajourhold av egne registre. Hvis eksisterende populasjon skal tildeles ny identifikator, vil det bli gitt tilbud om en slik oversettelsestjeneste eller andre tekniske løsninger som gjør det mulig å søke opp to ulike personnummer for en person.

7.1.5 Alternativ 4 – Alfamerisk kundennummer med åtte siffer

Oppbygging

Dette alternativet har mange likhetstrekk med Alternativ 3, med den forskjell at antall siffer er redusert fra 11 til 8 og at man kan benytte alfanumeriske tegn. Alternativet har følgende oppbygging:

$A_1A_2A_3A_4A_5A_6K_1K_2$

A: alfanumerisk individsymbol, K: kontrollsymbol.

Dette er en informasjonsløs identifikator, hvor tildeling skjer tilfeldig for ikke å angi opplysninger om fødselsdato, tidspunkt for registrering etc.

I likhet med Alternativ 3b, er det aktuelt å skifte identifikator for hele befolkningen.

Alternativet har to kontrollsifre.

Kapasitet og kontrollmekanisme

Dette alternativet har høy kapasitet, og denne øker med antall tillatte symboler. Alle symboler er tall eller bokstaver. Antallet symboler velges lik et primtall q , da dette er mest effektivt i forhold kapasitet og utregning av moduluskontroll. Det anbefales å benytte $q=23$, dvs. 23 symboler (tall og bokstaver) som innebærer at dagens modulo 11 endres til modulo 23. Under denne forutsetningen er den totale kapasiteten 148 millioner tilgjengelige identifikatorer, dvs. mer enn tilstrekkelig antall i flere hundre år framover. Det er også mulig å starte med et mindre alfabet, og utvide alfabetstørrelsen ved behov en gang i framtiden.

7.2 Robusthetsanalyse

Behovet for en ny personidentifikator med tilstrekkelig kapasitet fram til år 2150 er tiltakets prosjektutløsende behov. Kapasitet er omtalt både i behovsanalysen, målkapitlet og i kravkapitlet. Det er et effektmål at personidentifikatoren har robusthet ikke bare for dagens

populasjon, men også for eventuell framtidig utvidelse av populasjonen. Denne robusthetsanalysen av alternativenes kapasitet omfatter av den grunn to ulike utviklingsløp for framtidig personidentifikator:

- Personidentifikator omfatter dagens populasjon, dvs. at dagens kriterier for tildeling av fødselsnummer beholdes.
- Kapasitet for dagens populasjon og for de som i dag får D-nummer. Dagens kriterier for tildeling av fødselsnummer endres.

7.2.1 Kapasitet for dagens populasjon

Det vises til kapittel om krav, og avsnitt 5.1.1 hvor tilstrekkelig kapasitet for dagens registerpopulasjon minst fram til 2150 er et absolutt krav. Dette betyr at dersom et av alternativene ikke oppfyller dette kravet, vil det utelukkes som ny personidentifikator, og utgå fra videre analyse.

Tabellen nedenfor viser de ulike alternativenes kapasitet og om de er robuste fram til 2150.

Tidsperiode	0-alternativet Kapasitet pr. dag	Alternativ 1 Kapasitet pr. dag	Alternativ 2 Kapasitet pr. dag	Alternativ 3a og 3b Kapasitet totalt	Alternativ 4 Kapasitet totalt
2000-2039	496	826 per dato for hver verdi av kontrollsum K_1 (S_1 ulik 0). Total kapasitet 8260.	Minst 7535 per dato	123.8 mill. tilgjengelige identifikatorer (vasket register, 826.5 mill. før vask)	148 mill. tilgjengelige identifikatorer (for $q = 23$).
2040-2054 (eks. reserve)	331				
2040-2054 (inkl. reserve)	413				
2054-2150	For liten				

Tabell 10: Kapasiteter for alternativene

Analysen legger til grunn befolkningsprognoser for høy vekst fra Statistisk sentralbyrå fram til 2100, og en framskriving videre fram til 2150 (for utdypning se også kapittel 3.5 og vedlegg 1). Som det går fram av behovsanalysen er kravet til kapasitet minst 900 ulike verdier per dato for en informasjonsbærende identifikator og 37 millioner unike identifikatorverdier for en informasjonsløs identifikator. Dette er absolutte krav.

Som oversikten viser blir kapasiteten til dagens fødselsnummer for lav, og dette skjer med høy sannsynlighet fra 2040. I 2040 senkes kapasiteten fra 496 til 331 identifikatorverdier pr. dag (sum for begge kjønn). Riktignok er det store muligheter for at reserveserien står ubrukt slik at kapasiteten blir 413 identifikatorverdier pr. dag fra 2040. Også dette vurderer vi som for lav kapasitet. Dagens fødselsnummer vil ikke være robust identifikator hvor kravet til unik identifisering oppfylles fra og med 2040.

De to informasjonsbærende alternativene (Alternativ 1 og Alternativ 2) oppfyller kravet om minst 900 identifikatorverdier pr. dato. For Alternativ 1 er det tilstrekkelig å bruke to av 10 tilgjengelige nummerserier (sum for K_1 og K_2) for å registrere dagens populasjon. De to informasjonsløse identifikatorene oppfyller kravet om 37 millioner identifikatorverdier med god margin.

Det er ikke hensiktsmessig å rangere de fire alternativene til ny personidentifikator ut fra kapasitet. Alle fire har mer enn tilstrekkelig god kapasitet fram til år 2150.

7.2.2 Robusthet for utvidelse av dagens populasjon

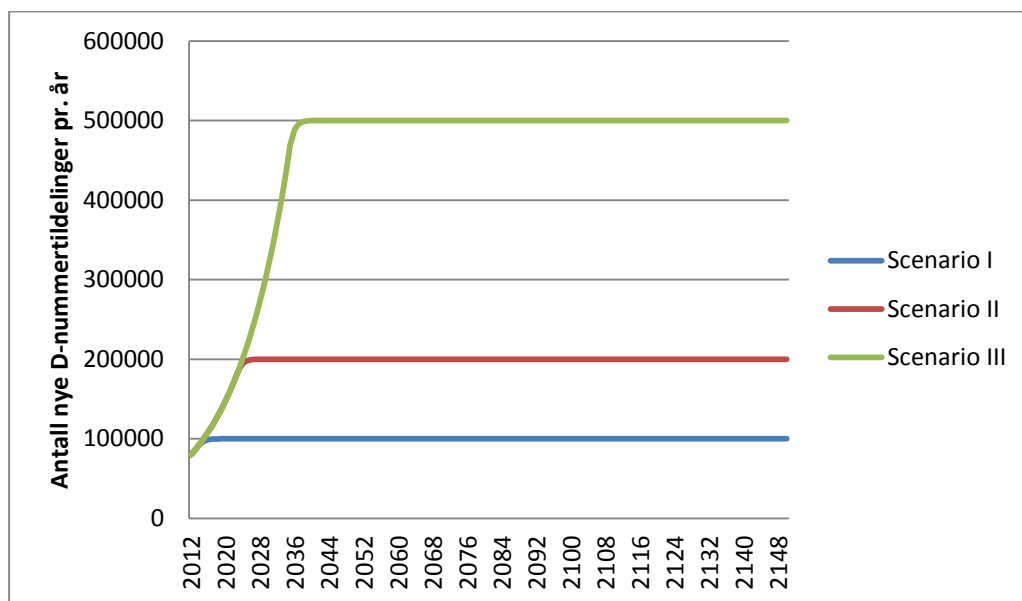
Under *Program for modernisering av Folkeregisteret* vurderes mulighetene for å endre dagens inndeling mellom fødselsnummer og D-nummer, som innebærer at man går bort fra dagens D-nummerordning. Det foreslås at en framtidig personidentifikator ikke inneholder informasjon som skiller mellom hvorvidt personen er bosatt eller har midlertidig tilknytning til landet. Informasjon om hvorvidt en person er bosatt vil framkomme som en statusopplysning i Folkeregisteret og gjøres tilgjengelig for brukerne av registeret (se avsnitt 3.2 om hvem som tildeles D-nummer i dag).

I vedlegg 3 er det gjennomført beregninger av hvor store kapasitetsbehovene er for en framtidig personidentifikator når både dagens populasjon for fødselsnummer og D-nummer skal inngå.

En rekke forhold skaper større usikkerhet når D-nummer blir inkludert i analysen. Etterspørselen etter D-nummer avhenger i stor grad av politiske og økonomiske forhold, situasjonen i Norge sammenliknet med andre europeiske land osv. På grunn av disse forhold har det vært vanskelig å utarbeide robuste prognoser for framtidig etterspørsel etter D-nummer. Beregningene gjennomføres i stedet med utgangspunkt i tre mulige scenarier for framtidig etterspørsel etter D-nummer. De tre scenariene er:

- Scenario I: 8 % årlig vekst i nye D-nr inntil man når 100 000 nye D-nummer pr år.
- Scenario II: 8 % årlig vekst i nye D-nr inntil man når 200 000 nye D-nummer pr år
- Scenario III: 8 % årlig vekst i nye D-nr inntil man når 500 000 nye D-nummer pr år.

I 2011 var tilveksten 80 000 nye D-nummer, med en vekst på 8 prosent fra 2010. Som det går fram av scenariene legger vi til grunn at en årlig vekst på 8 prosent vil stabilisere seg på henholdsvis 100 000, 200 000 og 500 000 nye D-nummer pr. år.



Figur 4: Scenarier for antall nye D-nummertildelinger pr. år $\tau(T)$,

For analysen har det vært hensiktsmessig å vurdere kapasitet i to forskjellige perspektiv:

Kapasitet i snever forstand

Det er ønskelig at individinformasjonen er korrekt for hvert enkelt individ. Dette forutsetter at antall tilgjengelige identifikatorverdier for hver enkelt dag og for hvert kjønn er tilstrekkelig stort til å dekke behovet. Det er derfor fornuftig å definere kapasitet i snever forstand som antall tilgjengelige identifikatorverdier pr. kjønn og dag. Kapasiteten i snever forstand tilfredsstiller behovet bare dersom det i populasjonen, for alle mulige fødselsdager og for hvert kjønn, ikke finnes flere individer enn denne kapasiteten.

Kapasitet i vid forstand

En kapasitet i snever forstand gir mening i en populasjon der antall individer fordeler seg tilfeldig men noenlunde jevnt over alle årets dager og over begge kjønn. Dette har tradisjonelt vært oppfylt for fødselsnumre. I den senere tid utgjør innvandring en stor del av nye tildelinger av fødselsnumre, og en konsekvens av dette er at populasjonen i mindre grad fordeler seg jevnt over alle datoer og begge kjønn. Det samme gjelder populasjonen som blir tildelt D-numre. Typisk vil behovet for identifikatorverdier være betydelig høyere for menn enn for kvinner, og mye høyere for datoer som 1. januar, 1. juli, 31. desember, på grunn av at populasjonen inneholder personer fra kulturer som ikke har tradisjon for eksakt registrering av fødselsdato.

Dersom kapasiteten i snever forstand i en slik populasjon skal oppfylle behovet, må alle datoer og begge kjønn ha «stor nok» kapasitet. En løsning på dette problemet er at når et nytt individ skal tildeles en identifikator og det ikke finnes flere identifikatorer som samsvarer med individets kjønn og offisielle fødselsdato, blir det valgt en identifikatorverdi for annen «vedtaksdato» innenfor samme år (og for samme kjønn.).

Denne pragmatiske tilnærmingen er mulig hvis den totale tilveksten i populasjonen ikke overstiger den totale tilgangen på identifikatorverdier pr. år. Det gir derfor mening å definere kapasitet i vid forstand for en informasjonsbærende identifikator som antall identifikatorverdier pr. år pr. kjønn. Dersom tilveksten i populasjonen pr. år overstiger denne vid-forstand-kapasiteten, er vedtaksdatometoden ikke lenger mulig.

Analysens resultat

I tabellen nedenfor har vi lagt til grunn SSBs prognose «høy nasjonal vekst» for populasjonen til fødselsnummeret kombinert med de tre scenario for D-nummer forklart over, ref. Figur 4 . På grunn av manglende prognoser for befolkningsvekst utover 2100, er det foretatt en rimelig ekstrapolasjon for utviklingen i populasjonen til dagens fødselsnummer fram til 2150. Dette er grunnlaget for resultatene i tabellen nedenfor (vedlegg 3).

De informasjonsbærende Alternativ 1 og 2 tilfredsstiller kapasitetsbehovet i vid forstand for alle de tre ulike vekstscenario frem til 2150. Ser vi på Alternativ 1 og alternativ 2 i de kombinerte scenariene «Høy nasjonal vekst» og «Scenario I» og «Høy nasjonal vekst» og «Scenario II» oppfylles også kapasitetskravene⁸.

For det kombinerte scenario «Høy nasjonal vekst» og «Scenario III» vil det være nødvendig å bruke opp 9 av 10 tilgjengelige nummerserier for å dekke behovet frem til 2100 for Alternativ 1, og likeledes 21 av 23 tilgjengelige nummerserier for Alternativ 2. I dette scenariet vil det bli nødvendig å utvide kapasiteten etter år 2100. Eller man kan innføre et strengere regime for tildeling av fødselsdato til innvandrere med uklart fødselstidspunkt slik at man får en mer jevn fordeling av fødetidspunkt gjennom året. En mulig løsning for Alternativ 1 er at man løser opp⁹ det andre kontrollsifferet, K₂, og dermed får mangedoblet kapasitet (mot svakere kontrollrutiner).

De informasjonsløse alternativ 3 og alternativ 4 tilfredsstiller samtlige scenarier fram til 2150.

Tabellen nedenfor illustrerer resultatene for Alternativ 1 og Alternativ 2.

Ulike vekstscenario	Kapasitet i snever forstand	Kapasitet i vid forstand
Høy vekst F-nummer D-nr., Scenario I	Tilstrekkelig	Tilstrekkelig
Høy vekst F-nummer D-nr., Scenario II	Tilstrekkelig	Tilstrekkelig
Høy vekst F-nummer D-nr., scenario III	Ikke tilstrekkelig	Tilstrekkelig

Tabell 11: Viser robusthet for Alternativ 1 og 2 for ulike vekstscenarier.

7.3 Prosjektperiode

Et forhold av særlig betydning for alternativanalysen er prosjektperiode eller forberedelsesperiode før ny personidentifikator innføres. Dette er et forhold som påvirker samtlige alternativer.

⁸ For det kombinerte scenariet «Høy nasjonal vekst» og «Scenario II» vil det være nødvendig å bruke opp 4 av 10 tilgjengelige nummerserier for å dekke behovet frem til år 2100 for Alternativ 1, og 9 av 23 tilgjengelige nummerserier for alternativ 2. Dette er nummerserier som kan anvendes fram til 2140 om de starter å brukes i år 2040 (angir århundre fra starttidspunkt og 100 år fram i tid).

⁹ Tillater flere lovlige sjekksummer i Mod11-kontroll av andre kontrollsiffer

Vi definerer prosjektperiode som den tidsperioden fra oppbyggingen til ny personidentifikator er offentlig kjent til den dato hvor virksomheter skal være i stand til å utveksle personinformasjon ved hjelp av ny personidentifikator.

Hvis vi tar utgangspunkt i at ny identifikator besluttet innen 2015, vil det være maksimum 25 år fra beslutning til ny identifikator må være innført i år 2040. Dette er det lengst mulige tidsløpet virksomhetene har til å planlegge, utvikle og implementere løsninger hvor ny personidentifikator benyttes. For å gi rom for nødvendig slakk regner vi 20 år som det lengste realistiske prosjektperioden. I motsatt ende kan vi forutsette at 5 år er den kortest mulige tidsperioden, dvs. at virksomhetene blir gitt 5 år til å gjennomføre prosjektet med å innføre ny personidentifikator.

Det er argumenter både for og mot å ha en lang prosjektperiode.

På den ene siden vil en lang overgangsperiode gi mulighet for en naturlig oppgradering/utskifting av dagens IKT-systemer hvor det inkluderes løsning for bruk av ny personidentifikator. Ved å ha kjennskap til hva som må endres, kan disse i noen grad gjennomføres uten å foreta ekstraordinære investeringer. Endringene gjennomføres i forbindelse med naturlig oppgradering/utskifting av løsninger og utstyr. Dette betyr også at en lang overgangsordning/prosjektperiode kan redusere prosjektets kompleksitet og risiko.

På den andre siden kan en lang overgangsperiode gi utfordringer for prosjektet fordi det er krevende å holde oppmerksomheten og aktiviteten oppe i et prosjekt som har mange års varighet (for eksempel over 10 år). Når prosjekter har lang tidshorisont, er det en reell fare for at temaet ikke oppfattes som aktuelt og blir nedprioritert i virksomheten. Det vil si at aktivitetene utsettes til det tidspunkt hvor man får dårlig tid for innføringen. Personer med sentral kompetanse kan dessuten flytte på seg og kontinuiteten rammes. Dette betyr at en "for lang" overgangsordning/prosjektperiode også kan øke prosjektets kompleksitet og risiko.

7.4 Metode for den samfunnsøkonomiske analysen

En samfunnsøkonomisk analyse er en systematisk innhenting og sammenstilling av informasjon om fordeler (nytte/gevinster) og ulemper (kostnader) som grunnlag for å ta en beslutning. Analysen skal basere seg på mest mulig fullstendig og sammenlignbar informasjon. Den skal synliggjøre relevante virkninger før det eventuelt fattes en beslutning om å iverksette tiltaket.

Videre skal man i en samfunnsøkonomisk analyse ta med virkninger for alle berørte parter i samfunnet. Dette innebærer at man vurderer økonomiske virkninger utover virkninger for den statlige virksomheten selv, og tar med virkninger for brukerne og for alle andre i samfunnet som blir berørt av tiltaket.

I Finansdepartementets veileder for samfunnsøkonomiske analyser er det beskrevet tre hovedtyper av samfunnsøkonomiske analyser:

- **Nytte-/kostnadsanalyse:** En systematisk kartlegging av fordeler og ulemper ved et tiltak. Nyttevirkninger og kostnader verdsettes i kroner så langt det er faglig forsvarlig.
- **Kostnadseffektivitetsanalyse:** En systematisk verdsetting av kostnadene ved ulike alternative tiltak som kan nå samme mål. Kostnadene verdsettes i kroner, og man søker å finne den rimeligste måten å nå et gitt mål.
- **Kostnads-/virkningsanalyse:** En kartlegging av kostnader for ulike tiltak som er rettet mot samme problem, men der effektene av tiltakene ikke er helt like. En kan i slike tilfeller ikke uten videre velge det tiltaket som har lavest kostnader.

Dette er å anse som en hybrid mellom kostnadseffektivitetsanalyse og kostnads-/virkningsanalyse. Man hadde ambisjoner om en nytte-/kostnadsanalyse, og det har vært

ramme for utforming av spørreskjema og datainnsamling. Datainnsamlingen har imidlertid vist at det er svært få nytteeffekter utover å videreføre dagens personidentifikator, og at disse i alle tilfelle er vanskelige å prissette og uansett neglisjerbare som bestemmende for dette konseptvalget.

De nytte- og kostnadseffektene som ikke prissettes vurderes i henhold til rammeverket for vurdering av ikke-verdsatte/ikke-prissatte effekter (den såkalte pluss-minus metoden). En gjennomgang av de prissatte kostnadene er gjort i kapittel 7.5.3.

7.4.1 Datainnsamling

Dette er et prosjekt hvor estimering av kostnadene er kompleks, både fordi mange virksomheter er berørt, og fordi det dreier om endring av et stort antall IT-systemer. Det er estimert at over 1000 IT-systemer vil bli berørt av endret fødselsnummer for en enkelt virksomhet.

Data er samlet inn gjennom en prosess i flere steg:

1. Utvikling av spørreskjema

En arbeidsgruppe bestående av representanter fra 7 større virksomheter fra både privat og offentlig sektor, deltok i arbeidet med å utvikle spørreskjemaet. I tidlig fase var over 40 ulike alternativer for ny personidentifikator med i vurderingen. Det ble lagt vekt på å få fram informasjon om både kvalitative og kvantitative effekter, og om prosessen før, under og etter innføring av ny personidentifikator.

2. Første utsendelse av spørreskjema

Spørreskjema ble i første omgang sendt til elleve virksomheter. Alle disse virksomhetene ble fulgt opp ved intervju samtale. Med bakgrunn i erfaringene fra denne runden ble spørreskjemaet revidert, og antall konsepter redusert fra 8 til 4 (i tillegg til 0-alternativet).

3. Andre utsendelse av spørreskjema

Spørreskjemaet ble sendt til om lag 50 virksomheter, inklusiv de elleve fra første runde.

Spørreskjemaet følger som vedlegg 4. Utvalget av virksomheter ble foretatt ved å velge et bredt utvalg av statlige virksomheter, et representativt utvalg av kommuner og noen av de sentrale virksomhetene i privat sektor. For privat sektor er finansnæringen spesielt berørt, og våre data er konsentrert til denne bransjen. Vi forventet at også telekom-sektoren ble påvirket i særlig grad, men denne effekten var lavere enn forventet. For statlig sektor er utvalget bredere, med ambisjon om å ha med berørte virksomheter fra relevante sektorer.

7.4.2 Forutsetninger for den samfunnsøkonomiske analysen

Følgende forutsetninger er lagt til grunn for den samfunnsøkonomiske analysen:

1. Nullalternativet (dagens situasjon) danner referansen som de øvrige alternativer skal sammenlignes mot. Dette er en tenkt situasjon der man levde videre med dagens fødselsnummer, det er ikke tatt med hvilke kostnader en slik situasjon ville innebære.
2. Å etablere en ny personidentifikator er en tjeneste uten slitasje, med en verdi som like godt kan øke over tid etter hvert som nye anvendelsesområder tar den i bruk og den er den sentrale nøkkel for å identifisere personer. Da kan levetiden for personidentifikator være flere hundre år, avhengig av utviklingen på IT-området. Alle alternativene dekker minst perioden fra nå og frem til 2150 og i praksis lenge etter dette. Analyseperioden settes som 2012-2150.
3. Restverdi settes til null som følge av svært lang levetid/analyseperiode. Dette er diskutert i kapittel 7.10.3.

4. Prisenivå er medio 2012. Tallene er i utgangspunktet ikke neddiskontert, da gjennomføringstidspunktet er uavklart, dette er imidlertid regnet ut med et eksempel i kapittel 7.6.2. Kapittel 7.10.1 diskuter effekter av diskontering.
5. Usystematisk risiko/usikkerhet: Alle prissatte parametere knyttet til kapitalflyt skal defineres som stokastiske variabler med spredning rundt sannsynlig verdi. Kalkulasjonsrenten betraktes som en deterministisk størrelse.
6. Alle økonomiske størrelser er gitt eksklusive MVA.
7. Det beregnes en skattekostnad av investeringskostnader og andre kontantstrømmer via offentlige budsjetter på 20 % av netto finansieringsbehov. Finansieringsbehovet er netto virkning på offentlige budsjetter, som beregnes som summen av lønnskostnader (inkludert skatt og arbeidsgiveravgift og innsatsvarer (eksklusiv merverdiavgift). Det skal regnes netto skattevirkning der det tas hensyn til betalte skatter, avgifter og overføringer som følge av prosjektet. Det betyr for eksempel at selv om det benyttes arbeidslønn inkludert skatt og arbeidsgiveravgift som kalkulasjonspris i beregningene, vil skatte- og avgiftsbidraget fra denne arbeidskraften bli tatt hensyn til når netto offentlig finansieringsbehov beregnes.
8. For dette prosjektet er systematisk risiko gjenspeilet i kalkulasjonsrenten svært lav, tilnærmet lik null. Prosjektet er ikke følsomt for økonomiske konjunkturer eller makroøkonomiske forhold. En kalkulasjonsrente på 2,5 prosent er en risikofri rente. Det er vanlig å anvende en kalkulasjonsrente mellom 4 og 6 prosent for statlige prosjekter. Til dette prosjektet er det valgt en kalkulasjonsrente på 2,5 prosent fordi systematisk risiko er svært lav. Et mulig scenario er at ny teknologi medfører en helt ny løsning for personidentifikator, eksempelvis ved at biometri innføres uten at fødselsnummeret anvendes som algoritme. For å ta høyde for en slik risiko diskuteres effekten av andre diskonteringsrenter i sensitivitetsanalysen (kapittel 7.10).
9. Korrelasjoner er hensyntatt ved at det er brukt usikkerhetsdrivere som dekker hele kostnadsbildet.

7.5 Nytte- og kostnadsestimater i analysen

7.5.1 Merkostnader

Et sentralt prinsipp i en konseptvalgutredning er at man sammenligner effekter mot det som ellers ville vært tilfelle. Dette tiltaket vil ha en lang planleggings- og gjennomføringsperiode og i denne perioden vil det gjennomføres større og mindre tiltak i mange av virksomhetene. Til virksomhetene er det understreket at man skal se på de *tilleggs kostnadene* som tiltaket vil medføre, *utover* kostnader som uansett påløper.

Man kunne tenke seg at *tilleggs kostnadene* ville være svært begrensede, men dette viser seg å ikke være tilfelle. Det vil uansett, selv i det enkleste alternativet, være et stort behov for å analysere hva som faktisk vil påvirkes og hvilke konsekvenser dette vil ha. Videre vil det påløpe betydelige ressurser til test, både innad i og mellom virksomheter. I konverteringsalternativene er det dessuten en svært omfattende jobb å konvertere eksisterende data, og disse kostnadene blir i liten grad redusert gjennom andre investeringer.

7.5.2 Kostnadselementer

Tiltaket innebærer en omfattende investering. I kontakt med virksomhetene viser det seg at det stort sett er de samme kostnadselementene som går igjen på tvers av virksomhetene. Den

relative størrelsen på kostnadselementene varierer dog mellom alternativene, spesielt gjelder følgende:

- Analysedelen er en større andel av totalkostnaden for Alternativ 1
- Alternativ 2 og 3a er å anse som gjennomsnittlige
- 3a, 3b og 4 gir ekstrakostnader for systemer der kjønn og/eller fødselsdato utledes fra fødselsnummer
- Alternativ 2 og 4 gir ekstrakostnader for systemer som ikke er tilrettelagt for alfanumeriske verdier, for eksempel en del stormaskinsystemer
- Alternativ 3b og 4 (konverteringsalternativene) har en langt høyere andel kostnader som gjelder integrasjoner
- Alternativ 3b og 4 har også større risiko for tap av liv og helse (gjelder helsesektoren)
- Alle alternativene vil kreve omfattende testing

Vi vil også understreke at det er betydelige kostnadsforskjeller mellom virksomheter innenfor samme bransje, for eksempel i finansnæringen er det store forskjeller for Alternativ 2.

7.5.3 Inndeling i kostnadsgrupper

I usikkerhetsanalysen er kostnadene som standard gruppert sammen i kostnadsgrupper. Det er her valgt en tilnærming der man grupperer de virksomhetene som ligner på hverandre, da usikkerhetsbildet for disse er sammenlignbart. I tillegg er det to spesielle grupper, *Uspesifisert* og *Merkostnader til forvaltning, drift og vedlikehold*. Vi står da igjen med følgende kostnadsgrupper:

- Helsevesenet
 - Her er det hentet inn tall og innspill fra de fleste regionale helseforetakene, samt at Helsedirektoratet har gitt supplerende opplysninger for resten av sektoren.
- Kommunesektoren
 - Her er det hentet inn tall og innspill fra et mindre utvalg kommuner, samt at Kommunesektorens organisasjon (tidligere Kommunenes Sentralforbund) har gitt innspill. Disse innspillene, samt innspill fra systemleverandører og egne vurderinger, er brukt for å utlede totalkostnader for sektoren.
- Statlige virksomheter med personopplysninger som kjernevirksomhet
 - I denne gruppen ligger de statlige virksomhetene som bruker personopplysninger i sitt daglige virke, enten i saksbehandling og/eller i integrasjoner med andre virksomheter. Eksempler inkluderer Skatteetaten, NAV, UDI og SSB.
- Statlige virksomheter uten personopplysninger som kjernevirksomhet
 - Disse kjennetegnes ved at de i mindre grad enn den forrige gruppen bruker personopplysninger i saksbehandling og integrasjoner.
- Finansbransjen
 - Her har det vært kontakt med flere av de store aktørene innen bank, forsikring og betalingsinnkreving
- Programvareleverandører

- Det har vært kontakt med de fleste store leverandører av standardsystemer. Disse kostnadene er imidlertid hovedsaklig fanget opp i de andre kostnadsgruppene, spesielt kommunesektoren.
- Uspesifisert
 - Denne er det verdt å dvele ved da den har et helt annet innhold enn det som er vanlig ved kostnadsestimering. Uspesifisert i denne sammenhengen inkluderer primært kostnader i andre virksomheter enn de som er fanget av øvrige kostnadsgrupper, for eksempel gjelder dette alle andre private virksomheter enn de som er nevnt over. Også det offentlige har kostnader under denne posten, deriblant en liten andel som gjelder fylkeskommunene.

Det har vært kontakt med mange virksomheter, men disse utgjør en liten del av totalpopulasjonen og usikkerheten er dermed betydelig. Uspesifisert har også et innslag av uspesifiserte virksomheter som kunne inngått i de andre gruppene.

Til slutt ligger det også i uspesifisert at noen av de foregående gruppene kan overlappe. Det er derfor en mulighet for at noen kostnader kan være talt dobbelt eller utelatt i begge gruppene. For eksempel gjelder dette mellom helsevesenet og kommunesektoren. Dette bidrar til å øke usikkerhetsspennet.
- Merkostnader for forvaltning, drift og vedlikehold i prosjektperioden
 - I og med at systemene som må erstattes har tilnærmet lik funksjonalitet som dagens, er det forutsatt at man på lang sikt vil ha de samme kostnader til forvaltning, drift og vedlikehold (FDV) som man ellers ville hatt. Men, i en større systemoppgradering /-utskifting vil enkelte systemer måtte kjøres i parallell og dermed påløper merkostnader for en periode. Noen virksomheter har gitt innspill til dette, spesielt på hvilke alternativer som påvirkes mest.

7.5.4 Fordeling av kostnader mellom sektorer

Tabellen nedenfor viser hvor stor kostnad som gjelder i de ulike sektorer. I feltene vises grunnkalkylene for kostnader, uten merkostnader til forvaltning, drift og vedlikehold i prosjektperioden.

Sektor	Alternativ 1	Alternativ 2	Alternativ 3a	Alternativ 3b	Alternativ 4
Kommuner	280	350	1 085	5 420	5 420
Helsevesenet	900	1 740	1 920	4 320	5 040
Andre offentlige virksomheter	565	1 475	1 210	2 650	2 960
Private virksomheter	420	2 115	1 155	4 335	5 400
Totalt	2 160	5 680	5 370	16 730	18 820

Tabell 12: Kostnader fordelt på sektor, tall i MNOK. Tall uten usikkerhet (grunnkalkyle), tillagt skattekostnad for offentlige utgifter. Ikke inkludert merkostnader til forvaltning, drift og vedlikehold i prosjektperioden.

7.5.5 Nytte

Den største nytteeffekten – og grunnen til at man setter i gang tiltaket – er at dette sikrer at man fortsatt har en funksjonell personidentifikator. Sammenlignet med en situasjon uten vil dette gi betydelige nytteeffekter som uten videre antas å overstige kostnadene for samtlige alternativer. Denne nytten vil være lik for samtlige alternativer.

Øvrig nytte er fanget opp i ikke-prissatte effekter og er beskrevet i kapittel 7.7.

7.6 Usikkerhetsanalyse

7.6.1 Resultater fra usikkerhetsanalysen

Under følger resultater fra usikkerhetsanalysen.

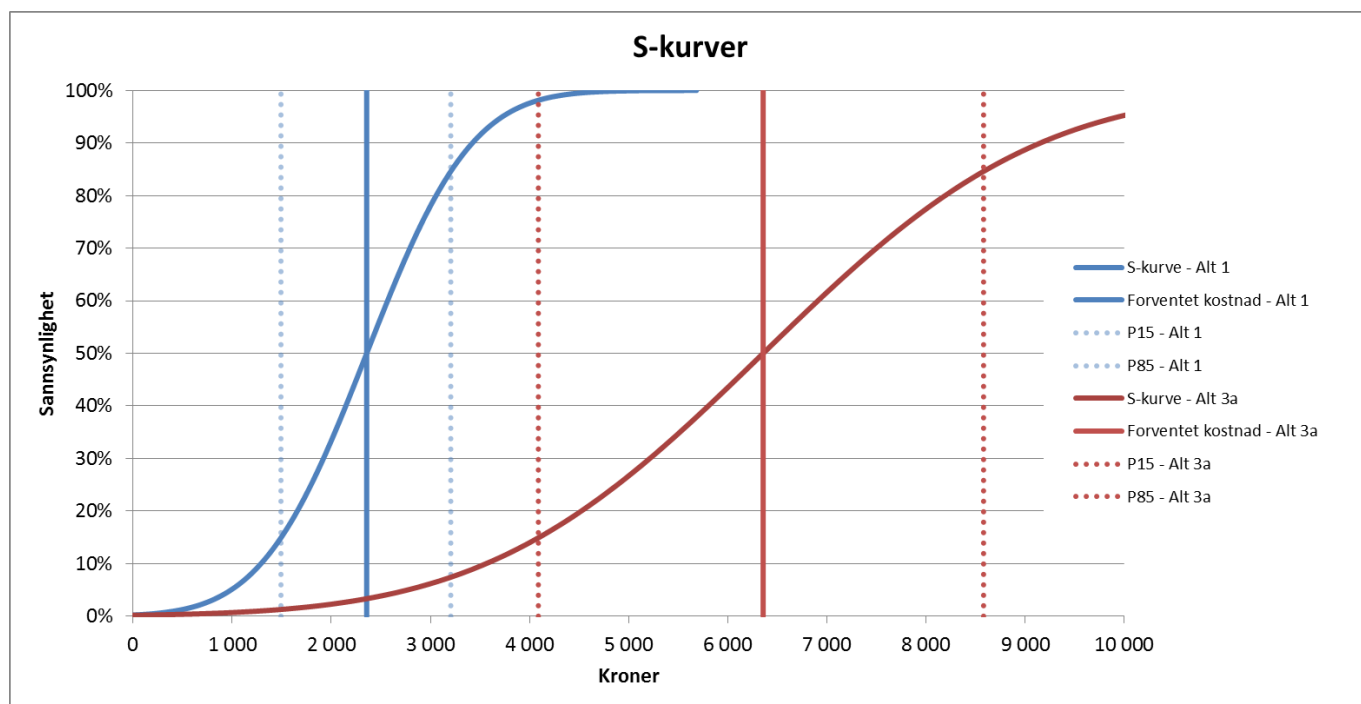
	Alternativ 1	Alternativ 2	Alternativ 3a	Alternativ 3b	Alternativ 4
Grunnkalkyle (uten usikkerhet)	2 230	5 860	5 540	17 270	19 430
Forventet kostnad	2 360	7 210	6 360	21 250	24 650
Standardavvik	830	2 370	2 180	8 270	9 460
Standardavvik (%)	35,3 %	32,9 %	34,3 %	38,9 %	38,4 %

Tabell 13: Resultater fra usikkerhetsanalysen, tall i MNOK. Tallene er inkludert skattekostnad for den andelen som gjelder offentlige utgifter

To av alternativene, 3b og 4, har betydelig høyere kostnader og høy usikkerhet.

Tabellen viser at blant alternativene som er informasjonsbærende er Alternativ 1 betydelig bedre enn Alternativ 2. Kostnaden er langt lavere og usikkerheten i absolutte tall tilsvarende.

De mest aktuelle alternativene er dermed Alternativ 1 og 3a, disse har følgende s-kurver.



Figur 5: S-kurver for Alternativ 1 og 3a

S-kurvene viser sannsynligheten for å overskride kostnadsnivåer. For eksempel viser kurven for Alternativ 3a at man med 85 % (100-15) sannsynlighet vil overskride ca. 4 000 MNOK i kostnader.

7.6.2 Effekt av periodisering og diskontering

Det er på nåværende tidspunkt stor usikkerhet knyttet til innføringstidspunkt.

En utregning av nåverdi (2012-kroner) er vist under. Denne har følgende forutsetninger:

- Innføringstidspunkt settes til 1. januar 2025.
- Diskonteringsrente er 2,5 %.
- Restverdi settes til null ved analyseperiodens slutt (2150).
- Kostnadene er periodisert til å fordele seg likt de fire siste årene før innføringstidspunktet. Dette er å regne som et vektet gjennomsnitt for samfunnet som helhet.

	Alternativ 1	Alternativ 2	Alternativ 3a	Alternativ 3b	Alternativ 4
Forventet kostnad	2 360	7 210	6 360	21 250	24 650
Netto nåverdi av kostnader	1 780	5 430	4 790	16 010	18 560

Tabell 14: Nåverdier for kostnadene til alternativene, tall i MNOK. Tallene er inkludert skattekostnad for den andelen som gjelder offentlige utgifter. Forutsetninger som beskrevet over

Med like forutsetninger ser vi at effekten er proporsjonal mellom alternativene og dermed ikke vil endre rangering eller vurdering av alternativene. Man kunne tenke seg at mer kompliserte

alternativer ville hatt lengre gjennomføringstid, men dette ville sannsynligvis aktualisert et senere innføringstidspunkt. Med tanke på nåverdi vil disse to effektene virke mot hverandre og resultatet påvirkes i liten grad.

7.7 Ikke prissatte effekter

De nytteeffekter som ikke prissettes vurderes i henhold til rammeverket for vurdering av ikke-verdsatte/ikke-prissatte effekter (pluss-minus metoden). Analysen følger veileder i samfunnsøkonomiske analyser, inkludert rundskriv om kalkulasjonsrente (jfr./32/ og /33/) og krav knyttet til Finansdepartementets KS-ordning.

Tre viktige begreper for vurdering av de ikke-prissatte effektene er: Betydning, omfang og konsekvens. Hvordan de tre begrepene anvendes er forklart nedenfor.

I det første trinnet vurderes betydningen tiltaket antas å ha for samfunnet. I stedet for å sette et tallanslag, etablerer man et gitt antall kvalitative kategorier fra liten til stor betydning, og plasserer deretter det aktuelle objektet i riktig kategori. Vi har benyttet kategoriene: Liten – middels – stor.

Neste trinn i metoden er å vurdere graden av disse endringene, dvs. omfanget tiltaket antas å skape for samfunnsutviklingen. Vi har benyttet fem kategorier for betydning: Stort negativt omfang, negativt omfang, intet omfang, positivt omfang og stort positivt omfang.

Det siste trinnet er konsekvens, som her betyr tiltakets virkninger for samfunnet sammenliknet med nullalternativet.

Vi kan oppsummere trinnene med at det gjennomføres en kvalitativ vurdering av betydning og omfang = virkning/konsekvens. Konsekvensene for samfunnet anslås langs en 9-delt skala fra (- - - -) til (+++++) som illustrert nedenfor.

++++	Meget stor positiv
+++	Stor positiv
++	Middels positiv
+	Liten positiv
0	Ubetydelig
-	Liten negativ
--	Middels negativ
---	Stor negativ
----	Meget stor negativ

Vurderingen av de ikke-prissatte nytteeffektene er gjort ut i fra innsamlet datamateriale (spørreskjemaundersøkelse), innspill fra referansegruppen og ut i fra diskusjoner på møter med interessentene.

Med bakgrunn i behovsanalysen, prosjektets mål og krav har vi identifisert følgende ikke prissatte effekter:

1. Ivareta et godt personvern
2. Brukervennlighet
3. Sannsynligheten for feilregistrering

4. Robusthet mot endringer av personopplysninger
5. Sameksistens med dagens fødselsnummer og andre brukte nummerserier

7.7.1 Ivareta et godt personvern

Behandling av personopplysning i Folkeregisteret er omfattet av personopplysningsloven, jfr. paragraf 3. Lovens formål er

“å beskytte den enkelte mot at personvernet blir krenket gjennom behandling av personopplysninger”.

Videre heter det at

“loven skal bidra til at personopplysninger blir behandlet i samsvar med grunnleggende personvern hensyn, herunder behov for personlig integritet, privatlivets fred og tilstrekkelig kvalitet på opplysningene.”

For å diskutere spørsmålet om personvern legger vi til grunn at fødselsnummeret er allment tilgjengelig og at det ikke er vanskelig å få kjennskap til andre personers fødselsnummer. Det er en utbredt misforståelse at fødselsnummeret er hemmelig, og at dette gir adgang til andre opplysninger. Fødselsnummeret alene er aldri et bevis for at vedkommende er den personen som man oppgir å være (autentisering), og er ikke en sikkerhetsmekanisme for å gi tilgang til opplysninger som er beskyttet. I praksis kan dette likevel skje. Da oppstår feil bruk med bakgrunn i feil rutiner. Dette er forhold som ligger utenfor analysens mandat, og er følgelig ikke gjenstand for vurdering.

Personvern kan være et tema i forhold til at man ved hjelp av personidentifikatoren knytter sammen og samkjører personopplysninger fra flere databaser. Til dette er det viktig å være klar over at fødselsnummeret som nevnt kun anvendes som identifikator og ikke for autentisering. Ved rett bruk skal man ikke få beskyttede opplysninger om personer ved kun å oppgi fødselsnummeret.

Hvem som bør ha tilgang til opplysninger i Folkeregisteret og hvilke opplysninger som registreres, er heller ikke tema i denne analysen¹⁰. Det er kun den informasjon som kommuniseres gjennom selve identifikatoren som er relevant i denne sammenhengen.

Identifikatoren gir informasjon om personers kjønn, alder og tilknytning til Norge¹¹. Kjønn og status i forhold til bosted kan ikke regnes som informasjon av personvernmessig betydning. For alder er det i større grad en vurderingssak om informasjonen ikke bør spres til virksomheter og andre personer når fødselsnummeret oppgis. Personer har ulik grad av sensitivitet i forhold til å kommunisere alder. Det finnes sammenhenger hvor dette er uønsket for flere, samtidig som det ikke er et sterkt argument i Norge. Det kan være “lite kjekt” å oppgi alderen, men det er få argumenter for at alderen må holdes skjult som en nødvendighet, eller at konsekvensene er alvorlige.

Vurdering av alternativene

En informasjonsløs identifikator kan med en viss sannsynlighet oppfattes mindre kontroversiell å utlevere enn fødselsnummeret slik det er i dag. De to informasjonsløse identifikatorene kommer litt bedre ut enn Alternativ 1 og 2, i og med at det er forbundet med et visst ubehag for en del personer å oppgi egen alder samtidig som man oppgir fødselsnummeret. Det er også slik at fødselsnummeret er offentlig kjent og registrert i en rekke sammenhenger.

¹⁰ Dette temaet er gjenstand for behandling i Forstudierapporten (3).

¹¹ Det at man er tildelt et fødselsnummer innebærer normalt at man enten er født i Norge, født av norske foreldre eller har vært bosatt i landet i minimum 6 måneder, jfr kapittel 3.1.

Det kan oppleves enklere å ha en alminnelig kjent personidentifikator om denne ikke inneholder informasjonselementer (Alternativ 3 og 4).

7.7.2 Brukervennlig personidentifikator

For å vurdere om en personidentifikator er brukervennlig er tilnærmingen å se på virkningen for individer/personer. I denne sammenheng forstår vi sluttbrukere som personer i ulike roller som privatpersoner. Sluttbrukere kan for eksempel være kunder, klienter eller pasienter.

Sluttbrukerne har rettigheter, plikter eller andre interesser i forhold til offentlige og private virksomheter som benytter fødselsnummer. Figuren nedenfor er en oversikt over ulike roller til sluttbrukerne.

Brukervennlighet er vanskelig å definere. Det er et begrep som er basert på empiri i en populasjon med stadig skiftende digital kompetanse, og er vanskelig å måle. Vi definerer en brukervennlig identifikator som:

En personidentifikator som er lett å huske, og som gir riktig og effektiv identifikasjon.

Hvilke koder er vanskelig å huske

Forskere har undersøkt menneskers evne til å huske lange serier med tall. Denne forskningen omfatter lengden på tallrekkefølger, evnen til å huske disse og evnen til å gjengi dem riktig. I tillegg til å huske hele tallrekkefølgen må vi passe på å taste dem inn helt riktig og i riktig rekkefølge (vedlegg 5 s. 10).

Dersom personinformasjon av typen fødselsdato er inkludert vil man kunne huske tegnstrenger av større lengde. Norsk forskning på korttidsminne viser at evnen til å gjengi en tallrekke korrekt faller raskt med tallrekkenes lengde (11). 98 prosent husker fire tall korrekt, 52 prosent husker åtte tall korrekt, 19 prosent husker 10 tall korrekt og kun to prosent husker tolv tall korrekt (vedlegg 5).

Vi husker dårligere når vi blir eldre. I kritiske situasjoner under stress blir hukommelsen dårligere. Det samme gjelder for personer påvirket av alkohol, narkotika eller enkelte medikamenter.

Et annet brukerbehov er at personidentifikatoren er stabil over tid. Innbyggerne har forventninger om å ha det samme fødselsnummeret hele livet, og bruken er tilpasset denne forventningen.

At fødselsnummeret er enkelt å huske, gir entydig identifisering og er stabilt over tid er viktige behov fordi vi kan komme i tilspissede situasjoner, for eksempel når en er syk, skadet eller i sjokk. Da er det viktig at de som gir behandling får korrekt medisinsk informasjon slik at behandlingen blir best mulig¹². Det samme gjelder for å informere pårørende.

Vurdering av alternativene

Informasjonsbærende identifikator er enklere å huske fordi de seks første sifrene inneholder informasjon om fødselsdato. For 0-



¹²Dagens fødselsnummer er ikke alltid tilgjengelig for helsepersonell i nødsituasjoner, da personer ikke alltid bærer ID-papirer eller har et norsk fødselsnummer. Helsevesenet har derfor opprettet et eget hjelpenummersystem for å kunne håndtere individer som trenger øyeblikkelig helsehjelp, men ikke kan oppgi fødselsnummer.

alternativet, Alternativ 1 og Alternativ 2 er det kun 5 sifre eller bokstaver som er tilfeldig sammensatt. 5 sifre er et antall tegn som de fleste av oss blir kjent med etter hvert og husker under forutsetning av at det er stabilt over tid.

Alternativ 3 er et kundenummer som består av 11 tegn. Forskning viser at dette er en såpass lang tegnstreng at mange ikke vil huske dette. Dette alternativet har lavere brukervennlighet av den grunn.

Alternativ 4 består av 8 tegn, tilfeldig sammensatt av sifre og bokstaver. Dette alternativet vil være enklere å huske enn Alternativ 3, men vanskeligere enn alternativene 0 - 2. Allerede med 8 tegn er man i den øvre grense for hvor lang informasjonsstreng man kan regne med at grupper av innbyggere vil kunne huske.

Alternativene 3b og 4 kommer dårligere ut mht. brukervennlighet fordi hele befolkningen skifter personidentifikator. I avsnittet over er stabilitet over tid omtalt som en faktor av betydning for brukervennligheten.

7.7.3 Sannsynligheten for feilregistrering

Dagens fødselsnummer er den autoritative nøkkelen for utveksling av personopplysninger mellom virksomheter i Norge. I tillegg anvender mange virksomheter dette som nøkkel også internt i virksomheten, hvor man slår opp på fødselsnummeret for å få fram opplysninger om personer og saker. I dag foregår vesentlige deler av kommunikasjon og dokumentgjenkjenning digitalt. Dette har bidratt til økt effektivitet og bedre samhandling i og på tvers av virksomhetene.

Feilregistrering av fødselsnummeret er ett forhold som påvirker kvaliteten til den elektroniske samhandlingen mellom virksomhetene.

Registrering foregår ofte manuelt, gjerne ved håndskrift eller ved verbal overføring. Dette innebærer at det kan inntreffe feil. Når elektronisk kommunikasjon øker som kommunikasjonskanal vil antall feil bli redusert fordi det er færre manuelle registreringer av fødselsnummeret.

Digitaliseringen av offentlig forvaltning og i samfunnet for øvrig medfører at kontrollsifrene får mindre betydning. Det utvikles digitale selvbetjeningsløsninger, hvor brukerne selv taster inn fødselsnummeret. For selvbetjeningsløsninger er det fortsatt viktig å ha en god feilkontroll, som gjennom datasystemer varsler når et ugyldig fødselsnummer anvendes.

Med dagens digitale forvaltning vil en feilregistrering som ikke oppdages eller rettes opp med en gang kunne spre seg elektronisk mellom datasystemer. Noen virksomheter har hyppige og automatisk oppdateringer i forhold til Det sentrale folkeregister, mens andre virksomheter oppdaterer sine data mot Det sentrale folkeregister sjeldent. Det er virksomhetene selv som styrer hvor ofte oppdatering skjer. Fram til i dag har dette vært en betalingstjeneste med fare for at noen virksomheter oppdaterer egne data mot Folkeregisteret for sjeldent.

Når en feilregistrering har skjedd kan saksbehandling på feil grunnlag skje. Feilen kan være av bagatellmessig karakter, men i noen anvendelser kan feil ha alvorlige konsekvenser. Et eksempel på dette er helsevesenet. Det kan også være svært alvorlig om feil går utover personers rettigheter eller plikter. Enkeltpersoner kan miste tjenester de har lovmessig krav på, eller de kan la være å utføre oppgaver eller gi riktig informasjon, som i ytterste konsekvens skaper mislighold.

Nedenfor beskriver vi sannsynligheten for feilregistreringer gjennom kontrollrutinene til de ulike alternativene. Det vil si i hvilken grad datasystemene klarer å varsle om feilregistreringer.

Generelt gjelder følgende:

- Ved bruk av ett kontrollsiffer (symbol) kan vi, ved riktig valg av metode, oppdage alle enkle feil og alle ombyttinger av to posisjoner.

- Ved bruk av to kontrollsiffer (symbol) kan vi, ved riktig valg av metode, oppdage alle doble feil, og alle ombyttinger av to posisjoner. Denne egenskapen er ikke oppfylt i dagens fødselsnummer.
- Muligheter for feilkorreksjon: Ved tastefeil kan et dataprogram produsere en kort liste av sannsynlig korrekte fødselsnumre.
- Både feildeteksjons- og feilkorreksjonsegenskaper forbedres betydelig med økende antall kontrollsifre. Vi anser det imidlertid ikke naturlig for denne anvendelsen å anbefale mer enn to kontrollsifre (vedlegg 2).

Vurdering av alternativene

0-alternativet

Dagens fødselsnummer ble utviklet i en tid hvor registrering hovedsakelig skjedde manuelt offline. Det ble lagt stor vekt på å ha gode kontrollmekanismer, og man valgte derfor to kontrollsymboler. Til sammenlikning kan vi oppgi at bankenes KID-nummer og det svenske folkeregisteret har ett kontrollsiffer.

Alternativ 1

For dette alternativet benyttes "1,5" kontrollsiffer. Det første kontrollsifferet er modulus 11, men med noe ulik oppbygging i forhold til dagens løsning. Det andre kontrollsifferet K_2 er identisk med dagens løsning.

Alternativ 1 har marginalt dårligere feildeteksjonsegenskaper sammenliknet med dagens situasjon. Dette er illustrert gjennom kurvene i figuren nedenfor.

Alternativ 2

Teoretisk kan det argumenteres for at en alfanumerisk kombinasjon skaper flere feil enn en ren numerisk kombinasjon, fordi tastaturet blir større, flere tegn tas i bruk og avstanden mellom dem øker. Antallet symbolkonflikter øker med økt antall tegn. Dette slår inn både for Alternativ 2 og Alternativ 4.

Imidlertid har Alternativ 2 en svært god feilkontroll som bidrar til at nesten alle feilregistreringene oppdages. Totalt sett kommer derfor dette alternativet litt bedre ut enn alternativene over.

Alternativ 3

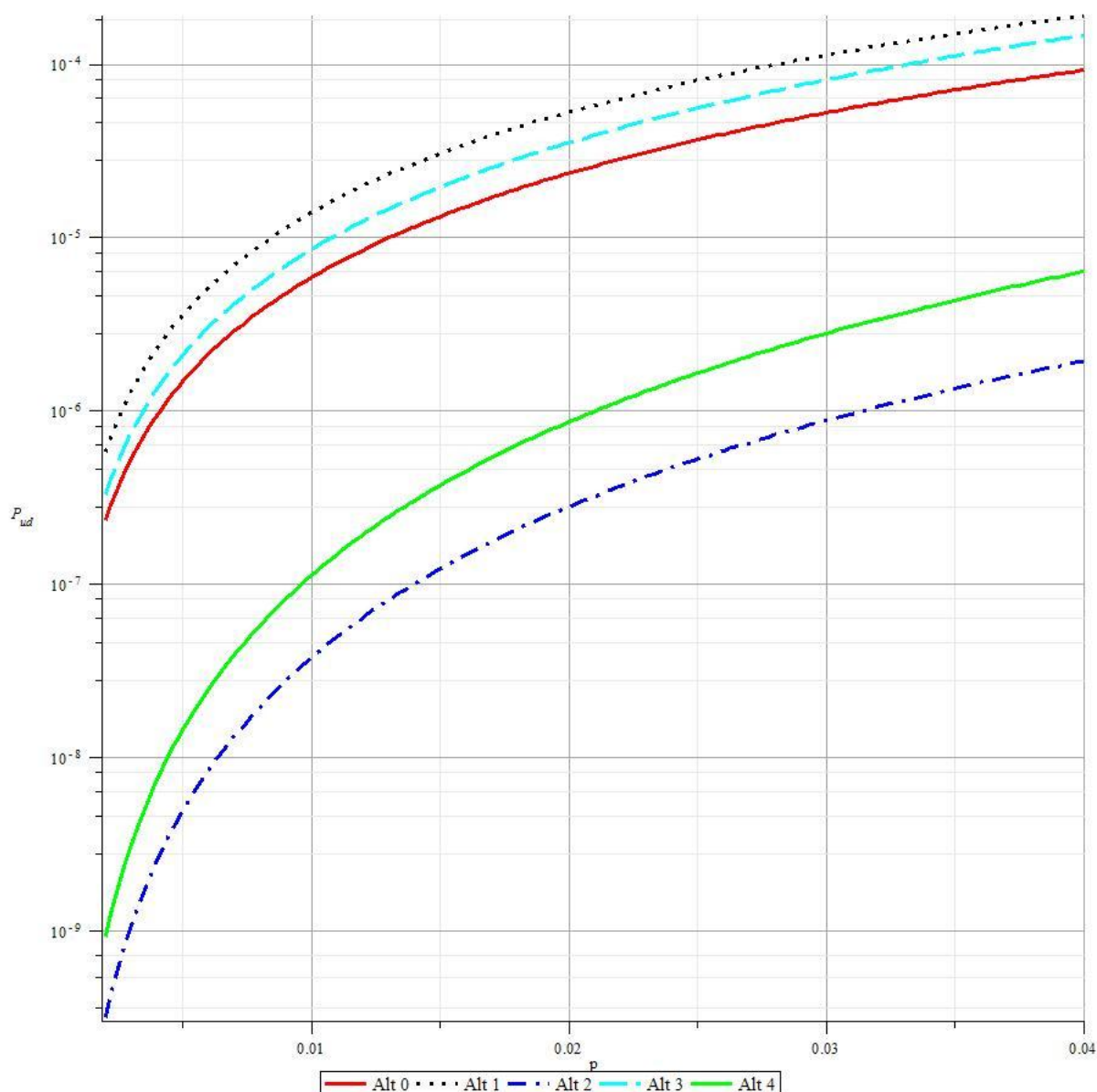
Alternativ 3 med 11 siffer, inkludert 2 kontrollsifre, og har feildeteksjonsegenskaper på lik linje med dagens fødselsnummer (marginalt dårligere).

Alternativ 4

Alternativ 4 er alfanumerisk og består av 8 tegn, med to kontrollsifre. Når antallet tegn reduseres fra 11 til 8 blir sannsynligheten for feilregistrering lavere, samtidig som det blir enklere for datasystemene å finne feil. Samtidig gjelder som for Alternativ 2, tastaturet er større og flere tegn øker sannsynligheten for tastefeil. Dette oppveier likevel ikke for den gode feilkontrollen. Dette alternativet kommer best ut sammen med Alternativ 2, og er bedre enn dagens løsning.

Sammenstilling

I figuren nedenfor har vi sammenstilt alternativenes sannsynlighet for uoppdagede feil. Utrekningene er foretatt ut fra alternativenes feildeteksjonsegenskaper (jfr. vedlegg 1). Det er Alternativ 2 og Alternativ 4 som kommer best ut i forhold til sannsynligheten for at feil i registrene ikke blir oppdaget. Alternativ 1 og Alternativ 3 kommer nokså likt ut som dagens løsning, nullalternativet.



Figur 6: Figuren viser sannsynligheten for uopptagede feil, P_{ud} som funksjon av p , der $p/(q-1)$ er sannsynligheten for at et symbol blir feiltolket som et av de andre $q-1$ lovligte tegnene i alfabetet og $(1-p)$ er sannsynligheten for feilfri tolkning av et enkelt symbol, ved bruk av henholdsvis Alternativ 0 - 4.

7.7.4 Robusthet mot endringer av personopplysninger

I 2011 var det 870 endringer av fødselsnummer i Folkeregisteret. Ny fødselsdato, endring av kjønn og feilregistreringer er de vanligste årsakene til endringer av fødselsnummeret. Feilregistreringer er behandlet i avsnittet over. I denne sammenheng er det endringer i de informasjonsbærende elementene, kjønn og fødselsdato som er tema.

Hyppigst forekommende er endring av fødselsdato blant innvandrere som ble registrert med feil fødselsdato (ofte ukjent fødselsdato) ved første gangs registrering. Virksomheter som har mange innvandrere i sine registre er særlig sårbar for denne problematikken, eks. UDI, helsesektoren, politi og NAV.

I prinsippet kunne deler av denne effekten (selve arbeidsmengden i saksbehandling) vært prissatt, men effekten av dette ville vært liten sammenlignet med kostnadene til alternativene.

Dette er en type virkning som strengt tatt kunne vært beregnet under prissatte virkninger. Sammenlignet med kostnadene for systemendringer ville virkningene være svært små. Det vil si virkningene for de informasjonsløse alternativene ville ha neglisjerbare nytteeffekter.

Det er også mer tekniske, ikke kvantifiserbare begrunnelser for å ønske en robust identifikator med høy fleksibilitet. Nedenfor har vi listet opp ulike argumenter som primære interessenter har gitt uttrykk for:

- Identifikator uten informasjon er "god utviklingsskikk" innen systemutvikling.
- Det gir mulighet for å tildele identifikator uten å kjenne de informasjonselementene som inngår i identifikatoren. For eksempel kan nyfødte barn eller til og med ufødte liv i mors mage få personidentifikator. Det siste får økt betydning ettersom behandling og medisinske inngrep i økt omfang blir gjennomført før fødsel.
- Slik fødselsnummerets personkrets er definert i dag, må de som har et D-nummer og senere blir bosatt i Norge få utstedt et fødselsnummer. En informasjonsløs identifikator kan forenkle overgangen fra D-nummer til Fødselsnummer for personer med usikker fødselstidspunkt¹³.
- Ut fra et sikkerhetsperspektiv er det vel også tryggere med en informasjonsløs identifikator. Et problem i dag er at fødselsnummeret blir "misbrukt" til en form for autentisering. En ny identifikator skal ikke kunne misbrukes til slike formål, bare framstå som et felles kundennummer som sikrer en stabil primærnøkkel og koblingsnøkkel i IT-løsninger. Mens personlige ID-kort, kombinert med registerinformasjon skal ivareta hensyn til autentisering og personvern.
- Dersom det blir aktuelt med personregister på tvers av landegrensene en gang i framtiden, kan en informasjonsløs identifikator enklere kunne tilpasses slik bruk (dette temaet er behandlet i et eget vedlegg 5) Dette er her vurdert som en lite aktuell problemstilling.
- Folkeregisteret er definert som nasjonal felleskomponent, dvs. et grunndataregister på linje med Enhetsregisteret og Matrikkelen. Grunndataregistrene bør framstå etter felles prinsipper. Både Enhetsregisteret og Matrikkelen har informasjonsløse identer med unntak av adresser (planlagt innført/under innføring).

Vurdering av alternativene

Det er et klart skille mellom de identifikatorer som er informasjonsbærende og de som ikke er informasjonsbærende i forhold til identifikatorens fleksibilitet. Man kan også si at det er et mål for hvor stabil personidentifikatoren er for endringer i omgivelsene.

Ikke informasjonsbærende alternativer

En personidentifikator som er nøytralt bygd opp som et kundennummer (Alternativ 3 og Alternativ 4), er mindre følsom for endringer av kjønn eller fødselsdato, ergo trenger heller ikke personidentifikatoren å endres som følge av ny informasjon. Dette gir en rekke fordeler for utvikling av registre og tekniske forhold rundt dette (se liste over).

¹³ En slik utvidelse kan skje uavhengig av alternativene i denne analysen, d.v.s. at også en informasjonsbærende personidentifikator kan inkludere dagens D-nummer populasjon. For personer med uklar identitet (fødselsdag) er det nødvendig å bestemme fødselsdag før tildeling av identifikator, og at denne ikke endres siden. Eventuelt må personen få nytt fødselsnummer.

Informasjonsbærende alternativer

Situasjonen er motsatt for dagens fødselsnummer, for Alternativ 1 og for Alternativ 2, hvor ny personidentifikator må tildeles om endring av fødselsdato eller kjønn inntreffer.

7.7.5 Sameksistens med dagens fødselsnummer og andre brukte nummerserier

For at innbyggerne kan beholde sitt fødselsnummer er det nødvendig å konstruere ny identifikator slik at den kan implementeres i god sameksistens med den gamle. Det vil si:

- For en gitt tegnsekvens må det gå klart frem hvilket identifikatorsystem den tilhører.
- Dette kravet oppfylles selv om et begrenset antall feil inntreffer.

I tillegg er det ønskelig at den nye identifikatoren ikke får verdier som sammenfaller med andre nasjonale nummerserier av betydning. Det er identifisert følgende nummerserier: D-nummer, hjelpenummer (helsesektoren) og kundennummer brukt i banksektoren.

Opplagte løsninger på sameksistensproblemet (i hvert fall versjon A) innebærer å bruke forskjellige lengder eller disjunkte alfabeter på eksisterende og ny identifikator. Det vil likevel være behov for å vurdere hvordan sameksistens fungerer særlig i forhold de tilfeller hvor feil inntreffer.

Nedenfor følger en vurdering av hvordan de ulike alternativene fungerer i sameksistens med andre nummerserier som eksisterer, og parallelt med dagens fødselsnummersystem.

Vurdering av alternativene

Alternativ 1

For dette alternativet vil de som har fødselsnummer i dagens ordning beholde dette. Når ny personidentifikator innføres eller det ikke er flere ledige identifikatorer, vil man skifte automatisk til nytt system for nye borgere/brukere. Brukerne er upåvirket av innføring av nytt system og trenger ikke vite hvilken type av fødselsnummer de har fått tildelt. Det oppstår ingen konfliktpunkter eller overlapp med andre typer nummerserier.

Alternativ 2

I likhet med Alternativ 1 vil de som har fødselsnummer i dagens ordning beholde dette. Det vil være lett å skille ny personidentifikator fra dagens fødselsnummer fordi den nye serien har minst tre bokstaver, som befinner seg på bestemte steder i rekken av symboler. Alternativ 2 er bygd opp på ett karakteristisk sett, som gjør at logikken er ulik andre nummerserier. Konflikt med andre nummerserier er med andre ord ikke et problem.

Alternativ 3

For Alternativ 3 er den totale mengden av mulige identifikatorer svært stor. Imidlertid er det en faktor at det eksisterer en rekke andre nummerserier med 11 tegn som personidentifikator av denne typen kan komme i konflikt med. Vi kan nevne dagens fødselsnummer-, D-nummerserien, hjelpenummer (helsesektoren) er eksempel på nasjonale nummerserier med 11 siffer. I tillegg er det en rekke andre virksomheter, for eksempel bankene, som anvender kundennummer med 11 siffer.

Det kan være behov for å "rydde bort" de viktigste nummertypene for å skape en ren ny nummerserie for dette alternativet. Dette betyr også at den høye kapasiteten som i utgangspunktet eksisterer for Alternativ 3, blir mindre avhengig av hvor "ren" man krever at den nye nummerserien skal være. Reserven av ubenyttede identifikatorer som er tilgjengelig ligger i mange tilfeller plundrete til, og det kan være uheldig og skape misforståelser dersom man bruker identifikator som er brukt i en annen sammenheng.

Når de viktigste nummerseriene er “vasket bort” (se vedlegg 3) har vi beregnet at Alternativ 3 har mer enn tilstrekkelig kapasitet (123 millioner identifikatorverdier). Likevel kan feil oppstå ved at man sammenblander nummer som i utgangspunktet har felles struktur.

Alternativ 4

Alternativ 4 er en informasjonsløs identifikator som er alfanumerisk og med åtte tegn. Formatet er så forskjellig (forskjellig alfabet og antall tegn) at det er svært lett å se forskjell på Alternativ 4 og dagens fødselsnummer. Dette alternativet har i liten grad overlapp med eksisterende nasjonale nummerserier.

7.7.6 Oppsummering av ikke-prissatte virkninger for de ulike alternativene

Ikke prissatt virkning	0-alternativet	Alternativ 1	Alternativ 2	Alternativ 3a	Alternativ 3b	Alternativ 4
Personvern	0	0	0	+	+	+
Brukervennlighet	0	0	-	---	----	---
Grad av feilregistreringer	0	0	+	0	0	+
Robusthet mot endringer av personopplysninger	0	0	0	+++	+++	+++
Sameksistens med andre nummerserier	0	0	+	--	--	+
Rangering	-	3	2	4	5	1

Tabell 15: Samlet oversikt over ikke-prissatte virkninger for alternativene

7.8 Vurdering av fordelingseffekter

For innbyggere som har uklar fødselsdato eller som velger å skifte kjønn, er det en ulempe med en informasjonsbærende identifikator. Særlig i gruppen innvandrere og asylsøkere finner vi enkeltpersoner med uklar fødselsdato som opplever at de må skifte fødselsnummer fordi fødselsdatoen endres. Med en informasjonsløs personidentifikator vil skifte av fødselsdato ikke skape problemer.

Gruppen av innvandrere og asylsøkere med uklart fødselstidspunkt består til dels av “ressurssvake” individer som opplever det ekstra belastende å skaffe seg nytt fødselsnummer. De har ofte språkproblemer og dårlig kjennskap til norsk offentlighet. Dette øker faren for feil i registreringen og mulighet for dobbeltregistrering.

Ved en konvertering av eksisterende fødselsnumre vil dette kunne være en tyngre prosess for svakerestilte, blant annet vil mange av disse ha tyngre for å lære seg nye numre og vite hva som må gjøres av den enkelte, for eksempel med tanke på identifikasjonspapirer.

Effektene over er dekket av ikke prissatte effekter og samlet sett vurderes fordelingseffekter å være av begrenset betydning som beslutningsrelevant faktor.

7.9 Vurdering av realopsjoner

7.9.1 Realopsjoner for nullalternativet

Man kunne tenke seg at ved å utsette beslutningen så fikk man et bedre beslutningsgrunnlag som gav helt andre utfall. Det er primært to situasjoner der dette kan tenkes:

- Det kunne tenkes en situasjon der man ikke lenger trengtes en unik personidentifikator. Dette kan for eksempel skyldes helt ny bruk av biometri. Dette regnes som så usannsynlig at denne realopsjonen har minimal verdi og kan ikke anbefales ettersom sjansen for manglende kapasitet øker fremover
- Den andre effekten er at ved å vente kunne flere virksomheter spare kostnader ved at datasystemer allerede er modernisert. Denne effekten er fanget opp i alternativene ved at man setter innføringstidspunkt langt frem i tid etter en beslutning. 10 år er antydnet fra mange virksomheter som optimalt, en lengre periode gjør at mange virksomheter legger arbeidet til side og mister kontinuitet

7.9.2 Realopsjoner mellom alternativene

Realopsjoner mellom alternativene begrenser seg til eventuelle utvidelser av kapasitetsbehovet. Dette er dekket i kapittel 7.2.

7.10 Sensitivitetsanalyse

7.10.1 Diskontering

Man kan velge å gjennomføre prosjektet på ulike tidspunkt og dette vil få innvirkning på nåverdien i dagens kroneverdi. Med en noenlunde lik lengde på prosjektgjennomføringen og uten verdsatte nytteeffekter vil dette imidlertid slå likt ut (i prosent) for samtlige alternativer og følgelig heller ikke påvirke rangeringen mellom dem.

7.10.2 Diskonteringsrente

Her gjelder samme resonnement som over, uten verdsatte nytteeffekter og med lik periodisering på prosjektgjennomføringen vil effekten av endret diskonteringsrente bli proporsjonal mellom alternativene og følgelig blir det ingen endring i rangeringen.

7.10.3 Restverdi

Det er i utgangspunktet ikke lagt inn noen restverdi på investeringen, da denne er satt opp med levetid helt til 2150.

Å etablere en ny personidentifikator er en tjeneste uten slitasje, med en verdi som like godt kan øke over tid etter hvert som nye anvendelsesområder tar den i bruk og den er den sentrale nøkkel for å identifisere personer. Da kan levetiden for personidentifikator være flere hundre år, avhengig av utviklingen på IT-området. Ut i fra dette resonnementet kunne restverdi settes lik investering. Det vil dog være usikkerhet knyttet til om helt andre teknologiske løsninger vil erstatte eksisterende løsning med personnummer.

Det er imidlertid liten grunn til at prinsippet for fastsetting av restverdien skulle variere mellom alternativene. Med en prosentvis sats, 100 % eller lavere, vil dermed effekten på alternativene

bli proporsjonalt lik. Størrelsen på denne effekten vil avhenge av analyseperiode og diskonteringsrente, med investering i 2025, analyseperiode til 2150 og 2,5 % rente første 40 år etterfulgt av 2 % rente de neste 85 vil restverdien være under syv prosent av investert beløp¹⁴ ved investeringstidspunkt. Med kortere analyseperiode og/eller lavere diskonteringsrente vil dette øke.

7.10.4 Størrelse på uspesifisert

En betydelig kostnadspost i analysen er uspesifisert, i denne sammenhengen virksomheter utover de som man har vært i kontakt med og/eller fått svar fra. Denne er satt som en prosent for øvrige kostnader, 25 % av beløp uten skattekostnad. Man kan tenke seg at denne prosentsatsen burde variere mellom alternativer, i så fall antas det at denne vil være høyere for de mer kompliserte alternativene – noe som vil øke de relative forskjellene mellom alternativene.

¹⁴ For en investering med normal risikoprofil og nye prinsipper for neddiskontering som foreslått i NOU 2012:16 vil dette være i størrelsesorden 2,5 % av investert beløp ved investeringstidspunkt.

7.11 Samlet samfunnsøkonomisk analyse

	Alt 0	Alt 1	Alt 2	Alt 3a	Alt 3b	Alt 4
Kostnader	-	2 360	7 210	6 360	21 250	24 650
Standard-avvik	-	35,3 %	32,9 %	34,3 %	38,9 %	38,4 %
Netto nåverdi av kostnader	-	1 780	5 430	4 790	16 010	18 560
Ikke-prissatte effekter (rangering)	-	3	2	4	5	1
Realopsjoner	Minimal realopsjon. Ikke tilstrekkelig kapasitet for utvidelse	Tilstrekkelig kapasitet for utvidelse	Tilstrekkelig kapasitet for utvidelse	Tilstrekkelig kapasitet for utvidelse	Tilstrekkelig kapasitet for utvidelse	Tilstrekkelig kapasitet for utvidelse
Fordelingseffekter	Som nå	Som nå	Som nå	Svakt bedre	Svakt dårligere	Svakt dårligere

Tabell 16: Samlet oversikt for alternativene

Nullalternativet er ikke aktuelt, da dette ikke oppfyller absolutte krav. Konverteringsalternativene (3b og 4) har svært høye kostnader og høy risiko. Selv om Alternativ 4 gjør det godt på ikke-prissatte effekter, er dette ikke i nærheten av å oppveie for høye kostnader og høy risiko.

Blant de informasjonsbærende alternativene er Alternativ 1 å foretrekke foran Alternativ 2. Kostnadene er langt lavere og forskjellene på ikke-prissatte effekter er små.

Alternativ 3a er eneste gjenværende alternativ som er informasjonsløst.

Samlet sett anbefales Alternativ 1 som har langt lavere kostnader enn Alternativ 3a. Også risikoen er lavest, målt i absolutte tall¹⁵. Dessuten er ikke-prissatte effekter noe bedre for alternativ 1.

Noen virksomheter foretrekker et informasjonsløst alternativ, i så fall er Alternativ 3a eneste realistiske mulighet. Jfr. kapittel 7.7.4 for argumenter for dette.

¹⁵ Som prosent er den noe høyere enn for alternativene 2 og 3a, dette som følge av at det er en prosent av et langt lavere tall.

7.12 Finansieringsplan

Finansieringen vil kunne avhenge av type virksomhet, størrelse på investering og andre faktorer. Utgangspunktet må være at virksomheter dekker utgifter over egne budsjetter og at vanlige prosesser for anskaffelser benyttes. Det antas at for enkelte offentlige virksomheter vil det kreves tilleggsbevilgninger, spesielt der implementeringen av ny personidentifikator gjøres sammen med andre IT-investeringer og/eller andre satsninger.

7.13 Hovedkonklusjon

Det anbefalte konseptet er Alternativ 1. Dette har langt lavere kostnader enn samtlige andre alternativer og risikoen er håndterbar. Alternativ 1 er bedre enn Alternativ 2 blant de informasjonsbærende, dette som følge av langt høyere kostnader for Alternativ 2 og kun beskjeden forbedring på ikke-prissatte effekter.

Alternativene med full konvertering av eksisterende fødselsnumre er svært kostbare og med svært høy risiko. I tillegg til risikoen uttrykt ved et høyt kostnadsmessig standardavvik har flere virksomheter stilt spørsmålstegn ved om prosjektet er gjennomførbart, dessuten er det varslet at det kan få konsekvenser for liv og helse.

Alternativ 3a, som ikke innebærer konvertering, er eneste gjenstående alternativ med informasjonsløs identifikator. Noen virksomheter foretrekker et slikt alternativ, men kostnaden er estimert til anslagsvis 4 milliarder kroner høyere og de samlede ikke-prissatte virkningene som marginalt dårligere enn Alternativ 1.

8 Føringer for forprosjektfasen

8.1 Innledning

Det er det nasjonale "fellesprosjektet" med å innføre ny personidentifikator som er gjenstand for omtale i dette kapitlet. Prosjekteiers muligheter til å lede og koordinere et nasjonalt prosjekt blir avgjørende for suksess. Det kan oppstå konkrete problemstillinger hvor det er viktig at prosjekteier prioriterer ut i fra fellesskapets interesser.

I tillegg til det nasjonale fellesprosjektet vil det pågå flere virksomhetsinterne prosjekter i berørte virksomheter. Det er ikke prosjekteiers ansvar å styre disse prosjektene, og vi har i dette kapitlet ikke vurdert hvordan de enkelte prosjektene i et utall virksomheter bør organiseres og styres.

Det er en betydelig iboende institusjonell usikkerhet ved at vesentlige deler av aktiviteten skjer fragmentert i et utall virksomheter, private og offentlige. Dette er en situasjon man ikke kan oppløse og som er utenfor prosjekteiers kontroll. Dette gir økt risiko. Sentrale grep blir informasjon, motivasjon og rammeverk.

Det forutsettes at valg av personidentifikator skjer med en bred administrativ og politisk forankring som støtter opp om prosjektet. Det anbefales at det endelige valget av personidentifikator gjøres "irreversibelt" i den grad dette er konstitusjonelt mulig.

Det vil være naturlig at Skattedirektoratet vurderes som prosjekteier.

Anbefalingene for forprosjektfasen er gitt med utgangspunkt i de vurderinger som er gjort i denne rapporten med vedlegg, og ut fra Finansdepartementets veileder med beskrivelse for innholdet i føringer for forprosjektfasen.

I forprosjektfasen foreslås utarbeidet Sentralt styringsdokument. Vi omtaler elementer fra KVUen som anbefales behandlet i Sentralt styringsdokument. Deretter er det omtale av suksessfaktorer og fallgruver, usikkerheter, gevinstrealiseringsplan og fagdepartementet som prosjekteier.

8.2 Viktige elementer fra analysen

Å innføre ny personidentifikator er et komplekst og utfordrende prosjekt av flere grunner. Det har direkte påvirkning på store deler av samfunnet, og et stort antall virksomheter er direkte avhengig av en godt fungerende personidentifikator. Et funksjonelt fødselsnummer eller tilsvarende løsning er samfunnskritisk og nødvendig for sikker utveksling og samhandling av personinformasjon mellom virksomhetene. Et stort antall virksomheter har fødselsnummeret som en sentral byggestein i egne datasystemer.

For et prosjekt med slik kompleksitet foreslås utarbeidet et sentralt styringsdokument for å støtte opp om prosjektgjennomføringen, og ikke minst for å lykkes i koordineringen mellom det store antallet virksomheter som blir direkte berørt. Dette innebærer bl.a. beskrivelse av:

- Overordnede rammer: Prosjektet skal ha en klar hensikt og entydige målsettinger, samt definerte rammebetingelser og grensesnitt.

- Prosjektstrategi: En klar strategi for hele forløpet slik at prosjektet kan gjennomføres på en sikker, effektiv og forutsigbar måte, herunder strategier for styring av usikkerhet, gjennomføring og organisering.
- Prosjektstyringsbasis: Beskrivelse av prosjektomfang inkl. prosjektnedbrytningsstruktur, kostnadsoverslag, finansieringsplan, leveranse-/tidsplan, organisering/styring og kvalitetsstyring. I forprosjektfasen vurderes grunnlaget for felles kontraktstrategi mellom virksomheter.
- Informasjons- og kommunikasjonsstrategi: Tiltak for å gi informasjon til interessentene og hvordan man skal involvere dem. Informasjon tilpasses og rettes inn mot ulike grupper av interessenter. Det etableres tett dialog med de mest sentrale interessentene i både plan- og gjennomføringsfasen. Et hovedprinsipp bør være at for mye informasjon er bedre enn for lite informasjon.

Styringsdokumentet bør bidra til avklaring, forankring og kommunikasjon mellom interessentene, samt skape felles forståelse for de oppgavene som skal løses.

Prosjektstyringsdokumentet utformes og detaljeres i samsvar med beslutninger fattet i *Program for modernisering av Folkeregisteret* med tilhørende prosjekter.

I kapitlene foran er det beskrevet en rekke forhold som det er behov for å bearbeide videre og konkretisere i forprosjektfasen. Nedenfor er kommentert enkelte områder som bør vies spesiell oppmerksomhet.

8.2.1 Behovsanalyse

I forprosjektfasen er det nødvendig å kartlegge interessentenes behov i forbindelse med prosjektgjennomføring og identifisere når viktige milepæler for prosjektet kan gjennomføres. Det er særlig viktig å vurdere prosjektets tidsplan og hvor lang tid interessentene trenger på å gjennomføre endringer i egen organisasjon. I KVUens behovsanalyse er dette gjort på et overordnet nivå med et utvalg interessenter. I en forprosjektfase forutsettes en sterkere involvering og forankring hos viktige interessenter. De ulike interessentenes roller identifiseres, og er grunnlag for videre involvering i prosjektet.

Konkret kan det nevnes at flere virksomheter har spilt inn at beslutning om ny personidentifikator med fordel kan tas raskt, av hensyn til pågående arbeid med å modernisere egne IT-systemer.

8.2.2 Kravdokument

Kravdokumentet er utformet som en del av KVUen og er derfor på et relativt overordnet nivå. I forprosjektfasen forutsettes at kravene blir bearbeidet videre og detaljert. Det vil være viktig å være tydelig på prioritering mellom ulike krav og prinsipper.

8.2.3 Gjennomføringsstrategi

For gjennomføringen av prosjektet er sentrale prioriteringer:

1. Understøtte de valg som gjøres for videre utvikling av Folkeregisteret. Gjennomføring av prosjektet bør sees i sammenheng med andre aktiviteter under *Program for modernisering av Folkeregisteret*. Et aktuelt tema er utvidelse av registerpopulasjonen til også å inkludere de grupper som i dag får D-nummer. Å avslutte dagens D-nummer ordning og utvide populasjonen til ny personidentifikator, kan legge føringer på prosjektets tidsløp. Det anbefales at dette avklares i løpet av forprosjektfasen.
2. God koordinering av det nasjonale prosjektet med vekt på standardisering og felles løsninger der dette er hensiktsmessig. Etablere overgangsordninger som reduserer kostnadene til innføring av ny personidentifikator.

3. Støtte virksomhetene i testing og annen samhandling med Folkeregisteret og virksomhet-til-virksomhet. Dette vil for en del virksomheter gjelde fra i svært tidlig fase. For innføring av ny personidentifikator er det kritisk at utveksling av informasjon mellom virksomhetene fungerer. Det er blant annet aktuelt å gjennomføre piloter.

Sikkerhet for liv og helse. Fødselsnummeret anvendes innen samfunnskritiske områder som helse, politi osv. hvor liv kan settes i fare om systemene ikke fungerer. Det er nødvendig å gjøre valg hvor man reduserer risiko for liv og helse.

I gjennomføringen av prosjektet bør det innarbeides styringsmessig fleksibilitet, uten at dette endrer premissene for ny personidentifikator som virksomhetene innretter seg etter. Tiltaket vil ha svært lang gjennomføringstid, noe som vil kunne medføre:

- Endrede rammebetingelser, eks. Folkeregisterets oppgaver endres.
- Teknologiske forutsetninger endres.
- Befolkningsutvikling og innvandring endres.

Før prosjektets oppstart må det skapes trygghet om at valget av personidentifikatoren er endelig. Skapes ikke dette oppstår en fare for at virksomhetene utsetter oppstart av eget arbeid i påvente av at premissendringer kan skje, med mulig konsekvens at arbeidet starter sent opp og kostnadene øker. Å ha en klart definert personidentifikator er nødvendig for virksomhetenes muligheter for en effektiv og god gjennomføring.

Uavhengig av hvilket alternativ som velges vil virksomhetene måtte håndtere to ulike identifikatorer i parallell. For Alternativ 1 vil forskjellene være små, og konsentrert til utregningen av moduluskontrollen av første kontrollsiffer (K_1). For de andre alternativene er det større forskjeller med mer krevende overgangsordninger.

Tidlig i prosjektet tas nødvendige beslutninger som gir hoveddrammene for prosjektet.

Nedenfor følger en utdyping av noen av de forhold som er viktig for at gjennomføringen skal lykkes.

8.3 Koordinert samhandling mellom virksomhetene

Det er viktig å kartlegge hvordan virksomhetsprosesser og samhandling mellom virksomhetene skjer med fødselsnummeret som nøkkel. Dersom det anbefalte Alternativ 1 velges, er det i prinsippet kun moduluskontrollen som endres. Det blir da viktig å identifisere hvilke systemer som anvender moduluskontroll og hvilke som ikke anvender dette. Folkeregisteret vil kunne understøtte denne jobben ved å utarbeide testdata på nytt format.

Det er behov for å etablere gode prosesser for dokumentasjon, informasjon og opplæring. Da kan flere virksomheter dra nytte av andres erfaringer.

Det bør vurderes å gjennomføre felles innkjøp av tjenester til offentlig sektor i forbindelse med innføring av ny personidentifikator. Dette kan for eksempel gjøres ved at det fremforhandles rammeavtaler som virksomheter kan gjøre avrop på. Dette er særlig aktuelt for områder med like systemer og felles utfordringer.

Det bør etableres tett samarbeid om testing mellom virksomhetene. Dette er en selvfølge i og med at utveksling av data mellom virksomhetene er en kritisk del. Det kan være aktuelt å utpeke en av de største og mest sentrale virksomhetene til å lede deler av test-arbeidet, og at de andre virksomhetene innretter seg etter denne virksomhetens tidsplan. Det er et poeng at når testing skal skje må de involverte virksomhetene være beredt.

8.4 Overgangsordninger

Overgangsordninger beskriver prosessen og etablering av tekniske verktøy knyttet til å innføre ny personidentifikator, dvs. hvordan man skal komme fra dagens situasjon og til en situasjon hvor innfasing av ny personidentifikator er slutført. Felles overgangsordninger vil kunne minimere kostnader og ulemper for virksomhetene.

Det foreslås at prosjekteier leder arbeidet med å planlegge, utvikle og implementere felles nasjonale løsninger og verktøy for innføringen. Det er aktuelt å etablere tekniske verktøy hos Folkeregisteret. Det kan være hensiktsmessig å utarbeide en felles implementeringsguide som et frivillig tilbud for virksomhetene. En slik guide kan inneholde felles retningslinjer, prinsipper, standarder, valideringsmekanismer med videre. Den kan også inneholde felles veiledning/løsning for å implementere ny modulusk kontroll. Gjennom dette kan prosjekteier bidra til at virksomhetene er samkjørte og at gjennomføringen skjer effektivt.

Det vil være flere slike fellestiltak som kan bidra til å gjøre implementeringen enklere for virksomhetene, øke sjansen for at man klarer å gjennomføre endringer tidsnok og redusere risikoen for at ting går galt.

Det er opp til den enkelte virksomhet å bestemme hvordan egen innfasing skal skje innenfor det definerte tidsrom. For eksempel hvor lang periode man velger til planlegging, utvikling og implementering av nye løsninger i forkant av at nytt fødselsnummer innføres. Dette kan sees på som tilsvarende som for en lov- eller regelendring, hvor virksomhetene må tilpasse seg den nye situasjonen og oppfylle nytt krav (nytt format) for elektronisk samhandling mellom virksomheter.

For mange virksomheter vil det være aktuelt å innføre ny personidentifikator i sammenheng med andre IT-investeringer og/eller andre prosjekter i egen virksomhet.

For enkelte sektorer er det aktuelt med samarbeid mellom virksomheter utover det prosjekteier kan tilby. For kommunene er tjenester som skal leveres til innbyggerne i stor grad de samme. En håndfull med standard IT-systemer er på markedet og svært mange kommuner anvender de samme systemene. Kunnskapsdeling, kontraktsamarbeid og koordinerte løsninger bør vurderes. Om kommunesektoren klarer å samarbeide er det muligheter for å ta ut betydelige stordriftsfordeler gjennom koordinert innføring, som vil redusere kostnadene og gi høyere kvalitet. Det kan være naturlig at KS/KommIT tar en koordinerende rolle for kommunesektoren. KommIT har selv spilt inn at det er potensial for samarbeid på maler (inkludert e-læring), forhandling med sentrale leverandører og noen grad av nye felleskomponenter, og vi støtter dem i dette.

Aktørene i den enkelte sektor bør selv ha ansvar for å utvikle sin sektors samarbeidsstrategi og tilpassede overgangsordninger for sektoren. Det bør vurderes om Folkeregistermyndigheten bør ha et apparat på plass for å gi støtte og råd ved behov.

8.5 Kontraktstrategi

Kontraktstrategien bør bygge på og ta utgangspunkt i tiltakets gjennomføringsstrategi. Generelt bør man gjennom en god kontraktstrategi kunne inngå kontrakter som beskriver ansvarsforhold, ressursinnsats, kompetanse og leveranser som både kunde og leverandør skal levere for å nå prosjektets mål. Det anbefales at kontraktene utformes slik at de er egnet til oppfølging av leverandør og godkjenning av leveranser i prosjektgjennomføringen. Det bør kunne skapes mekanismer slik at leverandørene har et reelt insitament for å bidra til prosjektets måloppnåelse.

Som utgangspunkt for virksomhetenes arbeid bør prosjekteier sentralt utarbeide et felles grunnlag som forenkler virksomhetenes arbeid med egen kontraktstrategi:

- Oppstilling av prosjektets behov og de krav som stilles – tekniske løsninger, tidsplan, mulige inndelinger, etc.
- Vurdering av grensesnitt, kompleksitet, størrelse og risiko.

8.6 Suksessfaktorer og fallgruver

Nedenfor er angitt noen sentrale suksessfaktorer og fallgruver som bør adresseres videre i forprosjektfasen.

8.6.1 Tidsløpet

I utgangspunktet har prosjektet tilstrekkelig fleksibilitet til å finne et optimalt tidspunkt for å slutføre implementering av ny personidentifikator, dvs. iverksettingsdatoen som er det tidspunkt hvor alle virksomheter skal være i stand til å håndtere ny personidentifikator.

Et langt tidsløp gir muligheter for naturlig oppgradering/utskifting av dagens IT-systemer hvor det inkluderes løsning for bruk av ny personidentifikator. Man har med andre ord kjennskap til hva som må endres, og kan gjennomføre dette i forbindelse med andre utskiftninger og oppgraderinger som finner sted. Slik naturlig utskifting/oppgradering reduserer kostnadene betydelig og kan i stor grad skje uten tilleggsinvesteringer for ny personidentifikator.

I forprosjektfasen eller tidligere må iverksettingsdato defineres, samt viktige milepæler som inkluderer flere virksomheter. Prosjektet har en dominoeffekt. Dersom en nøkkelvirkomhet ikke er klar til å innføre ny personidentifikator, påvirkes hele prosjektet og utsettelse er nødvendig. Tidsløpet bør settes med utgangspunkt i de tyngste virksomhetene/sektorene som trenger lengst tid. Innsamlet materiale tyder på at helsesektoren og kommunene er tunge sektorer i denne sammenheng. Helsesektoren fordi de har det største antall systemer med høy kompleksitet. Kommunesektoren fordi de er mange enheter med ulik tilgang av IT-ressurser. Også en del statlige virksomheter med stor bruk av fødselsnummer og grensesnitt mot andre virksomheter er viktige for tidsløpet.

Når tidsplan og iverksettingsdato er bestemt er det viktig at dette formidles bredt som absolutte datoer. Skapes det usikkerhet om dette er faren at mange virksomheter starter opp eget arbeid for seint.

I forprosjektfasen eller tidligere er det nødvendig å klargjøre mer detaljert:

- Optimal tidsperiode fra endelig valg av personidentifikator til iverksetting. Behovsanalysen angir om lag 10 år. Helsesektoren har dog spilt inn 20 år som ønsket prosjektperiode, og det er behov for å gå i dialog med helsesektoren om dette. Andre virksomheter har ønsket raskere prosjektgjennomføring på 2-3 år.
- Bestemme tidspunkt for iverksetting, hvor alle virksomheter er i stand til å behandle ny personidentifikator.
- Definere prosjektets faser og aktuelle milepæler for virksomhetene. Eksempel i hvilken periode testing for utveksling av informasjon mellom virksomhetene skjer.

Før iverksetting bør Folkeregisteret ha ferdigstilt sitt moderniseringsarbeid og gjennomført nødvendige endringer i god tid.

8.6.2 God eierstyring

Prosjekteier bør gjennomføre fellestiltak som forenkler virksomhetenes arbeid med å innføre ny personidentifikator. En viktig del av gjennomføringsstrategien er å definere opp ansvar og roller. Det bør gis klare føringer for:

- Hva som er prosjekteiers ansvar og oppgaver
- Hva som er virksomhetenes ansvar og oppgaver.

Særlig viktig er involvering og deltakelse fra de primære interessentene, som i sterkest grad er berørt av prosjektet.

Det blir en hovedutfordring å få virksomhetene til å engasjere seg i prosjektet. Lederforankring hos berørte virksomheter er nødvendig for å gi prosjektet prioritet og ressurser. Det foreslås at prosjekteier leder arbeidet med kommunikasjon og involvering.

8.6.3 Erfaringer fra tidligere prosjekter

For mange virksomheter er det aktuelt å se prosjektet i forhold til erfaringer fra år 2000-problematikken og fra telekom-sektorens endring av antall sifre i telefonnummeret. Selv om prosjektene er ulike har de også enkelte fellestrekk. Erfaringer er aktuelt i forhold til prosjektomfang, gjennomføring og organisering. En viktig forskjell er tid til rådighet. Denne gangen har virksomhetene lenger tid, og det er viktig at dette blir brukt til prosjektets fordel og for å redusere kostnadene. Endring av personidentifikator kan nå skje i forbindelse med systemendringer som likevel skal gjøres.

Det anbefales for øvrig at det i forprosjektet hentes inn erfaringer fra Sveriges innføring av justert personidentifikator.

8.7 Usikkerheter

Flere av de definerte usikkerhetene er utdypet og behandlet andre steder i dette kapittelet og spesielt under avsnitt 8.6, Suksessfaktorer og fallgruver.

Kombinasjonen av å ha et optimalt beslutningsløp og deretter en dyktig eierstyring med god koordinering og organisering fra prosjekteiers side kan få svært stor positiv effekt på kostnadene. Motsatt vil en dårlig prosess her kunne gi betydelig høyere kostnader og formodentlig også dårligere kvalitet.

Det er en betydelig usikkerhet knyttet til i hvilken grad kommunene klarer å samordne seg i sine investeringer. Her er det viktig å legge opp til den innsatsen som trengs for at man skal kunne komme i en fordelaktig situasjon. Når det gjelder helsevesenet vil et optimistisk utfall avhenge av at også andre prosesser på IT-området kommer riktig i gang, så her er det en betydelig restusikkerhet som ligger utenfor prosjektet.

Når det gjelder markedsusikkerheten er det begrenset hvilken påvirkningskraft man har på denne i de alternativene med de minste kostnadene, den vil avhenge av konjunkturer generelt og prisnivået på IT-investeringer spesielt. Dersom det skulle være aktuelt med et konverteringsalternativ måtte man imidlertid regnet med at størrelsen på prosjektet tilsier at markedet vil flyttes i negativ retning, det vil si at prisene øker fordi prosjektet vil ta en stor andel av den totale kapasiteten i bransjen.

8.8 Gevinstrealiseringsplan

En gevinstrealiseringsplan er en konkret og detaljert plan for å realisere de gevinstene som er identifisert. Med gevinst menes i denne sammenhengen gevinster i et samfunnsmessig perspektiv knyttet til de effektmålene som er satt for tiltaket, både de økonomiske og de kvalitative effektene.

Dette prosjektet har få/ingen angitte gevinster. Noen virksomheter oppgir at prosjektet vil utløse en gjennomgang og opprydding i IKT-systemporteføljen. Dette vil i neste omgang kunne ha positive effekter. Som analysen viser er det stor usikkerhet knyttet til slike virkninger, og at de samlet vil være små i forhold til kostnadene.

Det anbefales at det i forprosjektet klargjøres nærmere om gevinstpotensialet gir grunnlag for å utarbeide en gevinstrealiseringsplan, slik dette er beskrevet i veileder fra Finansdepartementet. Dersom prosjektet ikke har identifisert klare gevinster er det ikke grunnlag til å utarbeide en gevinstrealiseringsplan.

8.9 Prosjektorganisering

I denne sammenheng vil vi trekke fram følgende områder som aktuelt departement bør vie spesiell oppmerksomhet:

- Virksomhetenes rammevilkår
- Koordinering mellom departementene
- Styring og oppfølging

Virksomhetens rammevilkår

Det bør vurderes hva som er realistiske og gode rammer for implementering av ny personidentifikator. Det er for eksempel behov for å få en hensiktsmessig finansiering både til den som har det koordinerende ansvar, og til andre offentlige virksomheter som er berørt.

Koordinering mellom departementene

Det vil være et stort behov for koordinering og samarbeid mellom ulike statlige virksomheter, og det er naturlig at noe av dette understøttes ved at også departementene har en koordinering seg imellom.

Styring og oppfølging

For å sikre et vellykket prosjekresultat er det viktig med god og relevant styring og oppfølging. Det er behov for å gjennomgå roller og ansvar til berørte departement, og hvilke styrings- og rapporteringsrutiner de skal ha til koordinerende departement.

Rapportering til ansvarlig departement bør skje gjennom noen sentrale hovedmilepæler. Det bør vurderes hvilke behov for styring og oppfølging som eksisterer utover dette.

Bibliografi

1. **Grunndatarapporten.** *Utveksling av grunndata på personinformasjonsområdet.* s.l. : Finansdepartementet, 2007.
2. **Statistisk Sentralbyrå.** *Fødselsnummeret, oppbygging – kontrollsiffer, løsning etter år 2000.* s.l. : SSB.
3. **Skattedirektoratet.** *Modernisering av Folkeregisteret. Forstudierapport.* 2013.
4. —. *Ny distribusjons- og prosmodell for Folkeregisteret.* 2012.
5. **Finansdepartementet.** *Veileder nr. 11, Konseptvalg og detaljeringsgrad .* s.l. : Concept NTNU, 2010.
6. **Ytrehus, Helleseth Tor og Øyvind.** *Rapport om egenskaper ved ny personidentifikator, del 1.* s.l. : Seltersenteret, Universitetet i Bergen, 2012.
7. **Ytrehus, Tor Helleseth og Øyvind.** *Rapport om egenskaper ved ny personidentifikator. Del 2.* s.l. : Universitetet i Bergen, 2013.
8. **Skattedirektoratet.** *Strategi for modernisering av Folkeregisteret.* s.l. : Skattedirektoratet, 2011.
9. **Finansdepartementet.** *Veileder nr. 9 Utarbeidelse av KVU/KU dokumenter.* 2010.
10. *LOV 1970-01-16 nr. 01 Lov om folkeregistrering (folkeregisterloven), § 4.* 1970.
11. **Nordby, Knut.** *The Expanding Telephone number: Users' Needs for a Common Address Format in Future Converging Networks.* 2001.
12. **Finansdepartementet.** *Veileder i samfunnsøkonomiske analyser.* s.l. : Finansdepartementet.
13. **Agenda Utredning og utvikling AS.** *Økonomisk/administrative konsekvenser av gruppens forslag (Grunndatarapporten).* 2007.
14. **Skattedirektoratet/Difi.** *Personidentifikatorprosjektet. Forstudie: Innledende vurderinger.* 2011.
15. **Finansdepartementet.** *Rammeavtale mellom Finansdepartementet og Metier/Møreforskning Molde om kvalitetssikring av konseptvalg, samt styringsunderlag og kostnadsoverslag for valgt prosjekteralternativ. .* s.l. : Finansdepartementet, 2011.
16. **Skattedirektoratet.** *Menneskelige faktorer ved valg av ny personidentifikator.* 2013.
17. **Ytrehus, Tor helleseth og Øyvind.** *Rapport om D-nummer og økt populasjon for personidentifikator.* s.l. : Universitetet i Bergen/Seltersenteret, 2013.
18. **Finansdepartementet.** *Veileder nr. 9 Målstruktur og målformulering.* 2010.

Vedlegg

Rapport om egenskaper ved ny personidentifikator
Tor Hellesteth og Øyvind Ytrehus
Universitetet i Bergen/Selmersenteret
Revidert 04.04.2013

Innhold

1)	Konklusjoner og anbefalinger.....	2
2)	Bakgrunn og mandat	2
3)	Kriterier for vurdering av personidentifikator	3
3.1	Informasjonsløs identifikator kontra identifikator med dato og annen informasjon	4
3.2	Kapasitet.....	5
3.3	Feilkontroll	5
3.4	Lengde og huskbarhet/brukervennlighet	6
3.5	Alfabet og alfabetstørrelse.....	6
3.6	Kompatibilitet og sameksistens med dagens identifikator.....	7
3.7	Utvidbarhet.....	7
3.8	Entydighet	8
3.9	Programvareendringer	8
3.10	Teknologi, datasikkerhet og personvern	8
4)	Svar på oppgaver	8
4.1	Vurdering av kapasitet i eksisterende identifikator.....	8
4.2	Numeriske tegn kontra alfanumeriske tegn	10
4.3	Feilkontroll	11
4.4	Oversikt over alternativer.....	12
4.5	Praktisk gjennomføring.....	14
4.6	Andre alternativ	15
	Appendiks 1: Innføring i feildetekterende koder	15
	Appendiks 2: Beregning av kontrollsiffer i dagens fødselsnummer	18
	Appendiks 3: Oversikt over mulige identifikatorer	18
	Informasjonsbærende med numeriske tegn:	19
	Alt 0. - Nullalternativet. Dagens fødselsnummer - 11 posisjoner, 2 kontrollsifre	19
	Alt 0+: Dagens fødselsnummer - 11 posisjoner, 2 kontrollsifre	20
	Alt. 1. Dagens fødselsnummer med ett kontrollsiffer - 11 posisjoner	20
	Alt. 5. Skala for måned utvides	22
	Alt. 6. 14 sifre	22
	Informasjonsløs/begrenset informasjon samt numeriske tegn:	23
	Alt. 7. Dag inngår i individnummer, svensk løsning, 11 sifre og 2 kontrollsifre.	23
	Alt. 3. 11 sifre, 2 kontrollsifre	24
	Alfanumerisk og informasjonsbærende:	24
	Alt. 0X. 11 posisjoner, 2 kontrollsiffer. Tillater "siffer X" = 10	24
	Alt. 2. 11 posisjoner: Bruke alfabetutvidelse.....	25
	Alt. 8. 11 posisjoner, 2 kontrollsifre.	28
	Alfanumerisk og informasjonsløst/begrenset informasjon	28
	Alt. 4. Alfanumerisk, <i>m</i> informasjonsposisjoner	28
	Appendiks 4: Noen løsninger i andre land	30
	Appendiks 5: Krav til kapasitet	30
	Referanser	34

NB. Denne versjonen av rapporten er revidert i forhold til versjon datert april 2012, slik at alle vurderte alternativ er renummerert i henhold til senere dokumenter i saken.

1) Konklusjoner og anbefalinger

- Det er nødvendig å skifte ut eller utvide dagens fødselsnummer
- Anbefalinger, informasjonsløs identifikator:
 1. Alt. 4, med seks løpesymboler og to kontrollsymboler. 23 symbolers alfabet.
 2. Alt. 3, desimalt kundenummer. Ligner *litt for mye* på dagens system.
- Anbefalinger, identifikator med innbakt informasjon:
 1. Alt. 1 (Akseptabel løsning, liten forandring fra dagens system)
 2. Alt. 2, Variant 2 ("beste" løsning, større forandring i forhold til dagens system)

2) Bakgrunn og mandat

Dagens fødselsnummersystem ble etablert 1. oktober 1964 da oppbyggingen av et sentralt personregister startet. Alle som var bosatt i Norge på folketellingstidspunktet i 1960, fikk tildelt et eget nummer. Den eldste som fikk tildelt fødselsnummer var født i 1855. Alle bosatte etter 1964 har fortløpende fått tildelt fødselsnummer. Fødselsnummer tildeles i dag de som er bosatt i Norge i folkeregisterlovens forstand, de som er født i Norge samt norske statsborgere som aldri har bodd i Norge men som skal ha norsk pass.

Fødselsnummeret er på 11 siffer og har denne oppbyggingen:

$$D_1 D_2 M_1 M_2 \ddot{A}_1 \ddot{A}_2 I_1 I_2 I_3 K_1 K_2$$

der D: dato, M: måned, Å: år, I: individsiffer/personnummer, K: kontrollsiffer
I₃ viser kvinne (like tall) eller mann (ulike tall).

Første kontrollsiffer beregnes av de ni første sifrene i fødselsnummeret, mens det i utregningen av andre kontrollsiffer inkluderes også det første kontrollsifferet. For nærmere beskrivelse av beregningen av kontrollsymboler i dagens fødselsnummer, se "Appendiks 1: Innføring i feildetekterende koder" og "Appendiks 2: Beregning av kontrollsiffer i dagens fødselsnummer.

Mandat: Dagens fødselsnummerordning vil før eller siden komme opp i den situasjonen at den ikke vil ha kapasitet til å tildele entydige fødselsnumre til alle som trenger det. Skattedirektoratet har bedt oss (i) vurdere når denne situasjonen vil inntreffe og (ii) evaluere egenskaper ved nye identifikatorer for å utvide kapasiteten.

Konkret har vi forholdt oss til følgende oppdrag:

1. Verifisering og korreksjon av de vurderinger som er gjort i kapittel 2 [*i forespørselen*] angående egenskaper og kapasitet for dagens fødselsnummer.

2. *Hvilke konsekvenser (fordeler og ulemper) gir numeriske tegn kontra alfanumeriske tegn. Vil valg av tegntype ha betydning for moduluskontroll, se punkt 3 nedenfor.*
3. *Hvor mange kontrollsifre anbefales benyttet og med hvilken moduluskontroll. Hvis antall kontrollsifre og moduluskontroll anbefales forskjellig for ulike personidentifikatorer, skal analysen beskrive disse forskjeller med anbefaling for grupper av alternative personidentifikatorer, eventuelt for den enkelte personidentifikator.*
4. *Beregning av fremtidig kapasitet for de alternative nye personidentifikatorene som er beskrevet i kapittel 3 [i forespørselen]. Med kapasitet menes antall mulige unike identifikasjoner for den enkelte identifikator per dato. Her må det tas hensyn til eventuell bortfall pga. avvísninger i forbindelse med moduluskontroll etc. Hvor lenge vil den enkelte personidentifikator med dagens registerpopulasjon forventes å kunne benyttes med krav om unik identifisering, anvendelighet etc. Forklar hvordan dette praktisk kan gjennomføres.*
5. *Det er i kapittel 3 [i forespørselen] beskrevet åtte alternative nye personidentifikatorer. Hvis utreder ser andre mulige personidentifikatorer som anbefales utredet, bes det om at disse defineres og kort beskrives med hensyn til egenskaper etc.*

Oversikt over resten av denne rapporten: I avsnitt 3) vil vi diskutere forskjellige tekniske aspekter ved personidentifikatorer, og i hvilken grad disse kriteriene bør brukes for å velge blant alternative identifikatorer. Avsnitt 4) fokuserer på de konkrete oppgaver vi har fått. Appendiksene 1 og 2 beskriver henholdsvis feildetekterende koder generelt, og hvordan feildetekterende koder fungerer spesielt i dagens fødselsnummerordning. I Appendiks 3 går vi gjennom detaljerte beregninger for hver av de mulige identifikatorene vi har vurdert. Appendiks 4 inneholder en kort oversikt over hvilke løsninger som er valgt i noen andre land. Appendiks 5 diskuterer de kapasitetsbehovene vi har forholdt oss til.

3) Kriterier for vurdering av personidentifikator

En *personidentifikator* er en kort sekvens av tegn hentet fra et predefinert tegnssett eller alfabet, som skal brukes som et entydig ”navn” for å identifisere en person. I denne rapporten vil vi noen ganger bruke ”identifikator” i betydningen *en bestemt verdi* av en slik tegnstreng, mens vi andre ganger bruker ”identifikator” i betydningen ”system av identifikatorer”, dersom vi sammenligner slike systemer. Vi håper at det går fram av konteksten hvilken av disse betydningene vi refererer til. Dersom vi tror det er rom for tvil, vil vi bruke begrepene ”identifikatorverdi” og ”identifikatorsystem”.

I dette avsnittet vil vi diskutere forskjellige aspekt ved mulige identifikatorer, og i hvilken grad disse aspektene har betydning for valg av identifikator. De aspektene vi vil se på er:

1. Informasjonsløs identifikator kontra identifikator med dato og annen informasjon
2. Kapasitet
3. Feilkontroll
4. Lengde
5. Alfabetstørrelse og alfabet
6. Kompatibilitet og sameksistens med dagens identifikator
7. Utvidbarhet
8. Entydighet
9. Programvareendringer
10. Teknologi, Datasikkerhet og personvern

Nedenfor vil vi beskrive hvert av disse punktene mer i detalj.

3.1 Informasjonsløs identifikator kontra identifikator med dato og annen informasjon

Et prinsipielt spørsmål er om identifikatoren i seg selv skal inneholde *informasjon* eller ikke. Dagens fødselsnummer inneholder eksplisitt informasjon om fødselsdato (Dag, måned, to siste sifre i årstall), og i tillegg kodet informasjon om kjønn og århundre.

Praktiske aspekt: Dagens fødselsnummer er utviklet med hensyn på bestemte forutsetninger om bruken. En av disse er at registrering og bruk av identifikatoren ofte foregår offline, eller mer spesifikt uten en direkte kobling mot personregistre. Denne forutsetningen innebærer at blant annet at dersom en ønsker å bruke rudimentær informasjon om individet for enkel lokal prosessering uten oppslag i registre, må denne informasjonen legges direkte inn i identifikatoren.

Vi vil peke på at de institusjoner og bedrifter som har et legitimt behov for å utføre slik prosessering, i dag også har anledning til å innhente (fra offentlige registre) og lagre den aktuelle informasjonen, og at de så å si alltid har teknisk mulighet til å gjøre dette.

Hvordan bør informasjon kodes inn i en identifikator? Vi er av den oppfatning at dersom informasjon skal legges inn i identifikatoren, så bør denne informasjonen også kodes mest mulig tydelig, særlig for fødselsdato som er det viktigste informasjons-elementet i mange sammenhenger. Konsekvensen av dette er:

- Fødselsdato bør legges inn i ”klartekst”, for å gjøre den lettest mulig å lese, selv om det er teknisk mulig å finne mer kompakte representasjoner (se for eksempel alternativene 5 og 7 i Appendiks 3). Seks desimale sifre kan representere én million forskjellige verdier, men i hundre år er det (vanligvis!) 36576 forskjellige datoer på DDMMÅÅ-format. Det betyr at ”prisen” for å velge et klartekst-format kontra en kompakt representasjon er $6 - \log_{10}(36576) = 1,44$ desimalsymbol.
- Kjønn bør kodes slik at det er tilnærmet like mange identifikatorer for menn og kvinner. Det koster i utgangspunktet én bit å kode for kjønn; det vil si at mengden av identifikatorer deles i to omtrent like store undermengder. Samtidig finnes det antagelig *omtrent* like mange kvinner som menn for hver dag. Hvis det er *nøyaktig* like mange menn som kvinner for hver eneste dag, er det ingen kostnad forbundet med å kode for kjønn. Ettersom det vil være statistiske variasjoner fra dag til dag, og ettersom disse variasjonene vil være relativt større innenfor hver kjønnsgruppe enn i den samlede populasjonen, vil det likevel ha en viss kostnad å kode for kjønn – se Appendiks 5.
- Århundre, i den grad det er nødvendig, bør kodes på en enkel måte. Før man bestemmer hvordan århundre skal kodes, må man bestemme et antall k av ”gyldige” århundrer. Koding av århundre krever da minst $\log_{10}(k)$ desimalsymbol. I praksis betyr det at hvis det for hver DDMMÅÅ-dato er tilgjengelig et antall C gyldige løpenumre, så vil dette tallet reduseres til (i gjennomsnitt) høyst C/k når det skal kodes for århundre. For eksempel er dagens fødselsnummerordning essensielt tenkt for bruk over to århundre, fra 1855 til 2054. Som vi i viser i Avsnitt 4), er C lik 826 for dagens fødselsnummer, men på grunn av koding for århundre er det i gjennomsnitt ikke mer enn 413 nummer tilgjengelig for hver dag. Dagens fødselsnummer tildeles basert på århundre i henhold til en ad hoc-tabell. Alle nummer vil ikke bli utnyttet (se Tabell 1 og 2.)

Datasikkerhet og personvern. I noen land (se Appendiks 4) veier hensynet til *datasikkerhet* og *personvern* så tungt at man har valgt å ikke inkludere personinformasjon i identifikatoren, det vil si å bruke en informasjonsløs identifikator. Dette er også hovedanbefalingen i *Grunndatarapporten*.

En identifikator er *fullstendig* informasjonsløs bare dersom man ut fra å observere identifikatoren ikke kan trekke ut noen informasjon om individet. Dette forutsetter også at hvert individ blir tildelt en *tilfeldig* ledig identifikator (det vil si at identifikatorene ikke blir tildelt i en fast rekkefølge, fordi dette ville gi utenforstående en omtrentlig informasjon om tildelingstidspunkt.)

3.2 Kapasitet

Ethvert system for personidentifikatorer med fast lengde og alfabet vil ha et endelig antall forskjellige identifikatorer med en gitt egenskap. Dette antallet vil vi kalle *kapasiteten* til systemet. For informasjonsløse identifikatorer vil kapasiteten rett og slett se være det totale antall forskjellige identifikatorer.

For identifikatorer som inneholder eksplisitt informasjon, er det mer hensiktsmessig å uttrykke kapasitet med hensyn på hver spesifikk verdi av det eksplisitte informasjonsinnholdet. For eksempel bør man egentlig uttrykke kapasitet for dagens fødselsnummerordning i forhold til antall forskjellige tilgjengelige identifikatorverdier for hver verdi av (DDMMÅÅ, kjønn, århundre). I praksis vil man gjerne anta at det for hver fødselsdag *i gjennomsnitt* finnes like mange menn og kvinner, men kjønnsdeling av populasjonen vil likevel øke den relative variansen innenfor hver av de to gruppene sammenlignet med hele populasjonen – se Appendiks 5. Ut fra denne forutsetningen er det rimelig å definere kapasitet for dagens fødselsnummer som antall tilgjengelige løpenummerverdier for hver dato. For eksempel er beregninger av kapasiteten til dagens fødselsnummer beskrevet i avsnitt 4.1.

Det er ikke et naturlig krav at kapasiteten skal være så stor som mulig, fordi dette vil komme i konflikt med krav til andre parametre for identifikatorer. Derimot er det viktig at kapasiteten er *stor nok* i forhold til de prognosene som er tilgjengelig for behov.

Et alternativt, og av og til mer nyttig, mål på kapasitet er *varighet*. Dette er en verdi som fremkommer ved å koble kapasitetsmålet beskrevet ovenfor med (prognoser for) befolkningsvekst, og vil uttrykke hvor lenge et gitt identifikatorsystem kan brukes før man vil komme i en situasjon der det ikke er tilgjengelige identifikatorer (eventuelt med korrekt informasjonsinnhold) for tildeling. Dessverre er det vanskelig å lage pålitelige prognoser for hvor mange identifikatorer som er nødvendig for en gitt dag eller periode: dette tallet avhenger av (i) biologiske fødselsprosesser, som avhenger av blant annet befolkningsstørrelse og –alderssammensetning, og er vanskelig å forutse, (ii) innvandring, som er avhengig av globale økonomiske og politiske forhold, og er vanskelig å forutse, (iii) politiske beslutninger i forhold til hvilke grupper av individer som skal tildeles identifikatorer (bare statsborgere? Bosatte? Asylsøkere? Skattebetalere? Korttidsbesøkende?), som er umulig å forutse.

Vi har vurdert (avsnitt 4.1 og Appendiks 3) varighet i forhold til de prognoser vi har fått tilgang til fra Statistisk Sentralbyrå, som diskutert i Appendiks 5. På grunn av usikkerhet i prognosene, og dermed usikkerhet i anslag for varighet, vil vi imidlertid anbefale en konservativ holdning i forhold til kapasitet. Generelt er det relativt rimelig, i forhold til hvordan det påvirker andre parametre ved et identifikatorsystem, å sørge for at kapasiteten er ”robust” i forhold til ”marginal”.

3.3 Feilkontroll

Registrering foregår ofte manuelt, gjerne ved håndskrift, ved verbal overføring, og eventuelt ved hjelp av sviktende korttids- og langtidshukommelse. Dette innebærer at det ofte kan inntreffe feil. Disse feilene kan være av bagatellmessig karakter, men i noen anvendelser (for eksempel innen helsevesenet) kan feil ha katastrofale konsekvenser.

Som nevnt er dagens fødselsnummer utviklet med hensyn på en forutsetning om at registrering og bruk av identifikatoren ofte foregår offline, uten en direkte kobling mot personregistre, slik at kontroll av personopplysninger ikke er øyeblikkelig mulig eller tilgjengelig. I dag foregår registrering ofte, men ikke alltid, online. Med direkte tilgang til personregistre som vil gi supplerende personopplysninger og feilkontroll vil behovet for eksplisitt feilkontroll avta.

Ettersom online-registrering ikke alltid er mulig, vil vi likevel anbefale at man bruker en identifikator som inneholder kontrollsymboler, og som gjør det mulig å oppdage feil. I dagens fødselsnummer gir de to kontrollsymbolene mulighet til å oppdage alle enkle substitusjonsfeil (det vil si feil i enkle symbol), alle kombinasjoner av feil i to av de siste fem symboler, og noen feil som ifølge statistisk materiale forekommer oftere enn andre (se (Selmer)).

Dagens fødselsnummer gir ikke grunnlag for å oppdage alle kombinasjoner av feil i to vilkårlige posisjoner, men de systemene (med to kontrollsymboler) vi diskuterer i Appendix 3 vil oppnå dette – i tillegg til at de vil oppdage en del andre feil.

Kontrollsymboler gjør det i noen tilfeller mulig å utføre *feilkorreksjon*, dvs (i denne sammenhengen;) dersom et fødselsnummer er feil registrert kan et dataprogram produsere en kort liste av sannsynlig korrekte fødselsnumre.

Appendiks 1 og Appendix 2 gir henholdsvis en innføring i prinsipper for feildeteksjon, og en beskrivelse av hvordan disse prinsippene blir anvendt i dagens fødselsnummer.

3.4 Lengde og huskbarhet/brukervennlighet

Kapasitet kan lett oppnås ved å velge identifikatoren lang nok. I *databehandlingssammenheng* er det i dag ikke noe problem om identifikatoren er lang. Men en lang identifikator har åpenbare ulemper med hensyn på brukervennlighet:

- For eier av identifikator:
 - Å huske (særlig delen som ikke inneholder dato)
 - Å skrive ned
- For andre:
 - Å oppfatte, skriftlig eller verbalt
 - Å huske (korttidsminne)
 - Å skrive ned
- For de som skal lage papirskjema: Plasskrav

I tillegg øker sannsynlighet for feilregistrering med lengden på identifikator.

3.5 Alfabet og alfabetstørrelse

Med alfabetet $\mathcal{A}$ mener vi mengden av tillatte tegn i identifikatoren. For et gitt alfabet $\mathcal{A}$ definerer vi $s = s(\mathcal{A})$ til å være størrelsen til $\mathcal{A}$, det vil si antall tegn i $\mathcal{A}$. Mer generelt kan de forskjellige symbolposisjonene i identifikatoren ha forskjellige alfabeter.

I noen identifikatorer er det valgt å ha et numerisk og desimalt alfabet (for eksempel i det norske fødselsnummeret.) Dette er *antageligvis* basert på flere faktorer:

- Dato, som har en naturlig numerisk representasjon, er innbakt i identifikatoren,
- Kulturelle konvensjoner kan ha gjort det desimale alfabet til et naturlig valg for et ”identifikasjonsnummer”,
- I en periode kan det ha vært bedre tilgang til ”numerisk” tastatur enn til fullt tastatur (for eksempel QWERTY-tastatur).

Det er imidlertid ingen prinsipielle grunner til at alfabetet bør være desimalt for alle identifikatorposisjoner. Med hensyn på kapasitet vil det være en betydelig fordel å ha et stort alfabet, og med hensyn på feilkontroll (se Appendiks 1) vil det være en fordel å ha en alfabetstørrelse s ($\mathbb{Z}$) som er et primtall og som er større enn lengden på identifikatoren.

Det betyr ikke at det er uproblematisk å øke størrelsen på alfabetet. To symboler som kan forveksles (verbalt eller ved tolking av trykt eller håndskrevet tekst) bør ikke begge være lovlig symboler i alfabetet. For eksempel bør ikke tallsymbolet "0" og bokstaven "O" forekomme samtidig. (Dette er krav som et rent desimalt alfabet ikke ivaretar på en perfekt måte: Blant annet kan "3" ofte forveksles med "2", "5", eller "8".)

Når alfabetstørrelsen s er valgt, er valg av selve alfabetet i hovedsak uavhengig av andre designbeslutninger. Imidlertid kan uheldige valg av alfabet resultere i en forhøyet sannsynlighet for symbolfeil, noe som igjen har betydning for behov for feilkontroll.

3.6 Kompatibilitet og sameksistens med dagens identifikator

Det er ønskelig at dagens fødselsnummer skal bevares for de som har fått et tildelt. For å oppnå dette, må man konstruere ny identifikator slik at den kan implementeres i fredelig sameksistens med den gamle. Det vil si (A) at for en gitt tegnsekvens må det gå klart frem hvilket identifikatorsystem den tilhører, og (B) at dette kravet må oppfylles selv om et begrenset antall feil inntreffer.

I tillegg er det ønskelig at den nye identifikatoren ikke får verdier som sammenfaller med brukte eller gyldige D-Nummer, som kan oppfattes som elementer i samme nummerserie som fødselsnumrene. D-nummer er på formen

$$D_1^* D_2 M_1 M_2 \text{Å}_1 \text{Å}_2 I_1 I_2 I_3 K_1 K_2$$

der hvert symbol har samme betydning som i fødselsnummeret, bortsett fra at $D_1^* = D_1 + 4$, det vil si at det legges 40 til dagverdien i dato.

Opplagte løsninger på sameksistensproblemet (i hvert fall versjon A) innebærer å bruke forskjellige lengder eller disjunkte alfabeter på eksisterende og ny identifikator. Det vil likevel være behov for å vurdere hvordan sameksistens fungerer i versjon (B) av problemet, det vil si når feil inntreffer.

En løsning på (B) kan være å beholde lengde og utvide alfabet, men i det nye identifikatorsystemet å bare tillate identifikatorer som er "tilstrekkelig forskjellige" fra alle identifikatorverdier i det eksisterende identifikatorsystem. Alternativ 2 (se Appendiks 3) er et eksempel på en ny identifikator som er konstruert med henblikk på slik sameksistens.

3.7 Utvidbarhet

På samme måte som dagens fødselsnummer er i ferd med å bli oppbrukt, vil en ny identifikator med endelig kapasitet også en gang i fremtiden komme i samme situasjon. Da er det en fordel om identifikatoren er valgt slik at det er lett å utvide og slik at kompatibilitet med tidligere identifikatorer, som beskrevet i Avsnitt 3.6.

Det er relativt enkelt å utvide med hensyn på alfabetstørrelse og til en viss grad enkelt å utvide med hensyn på lengde. Derimot byr det ofte på problemer å utvide grunntallet, q , for modulusberegning av kontrollsymboler. Derfor vil en god strategi være å velge q som et primtall mye større enn den lengden som er aktuell i dag. Man kan likevel om ønskelig starte med et

mindre alfabet, selv om dette foreløpig begrenser kapasiteten. Når det blir nødvendig kan man øke alfabetet inntil det når størrelse q . Eksempler på slike utvidelser er diskutert i Appendiks 3), spesielt i forbindelse med Alternativ 4.

Det å ha kontrollsymboler innbakt i identifikatoren gir også en fremtidig mulighet til å utvide kapasitet, ved at kontrollsymbolet en gang i fremtiden kan brukes til å bytte ut pålitelighet mot økt kapasitet.

3.8 Entydighet

Dersom en gitt identifikatorstreng assosieres med to forskjellige individer, åpner det for forveksling av disse individene, og dette selvsagt svært uheldig. Dette forekommer for Alternativ 0+ (Appendiks 3). Ettersom Alternativ 0+ ikke gir en robust økning av kapasiteten, og ettersom det ikke er vanskelig å oppnå både entydighet og robust kapasitetsøkning på andre måter, anbefaler vi ikke løsninger som ikke tilfredsstiller entydighetskravet.

3.9 Programvareendringer

Alle som bruker dagens fødselsnummer og som har programvare som validerer denne identifikatoren ved å sjekke kontrollsymboler og datoformat, vil måtte oppdatere denne programvaren ved en endring av formatet – det vil si, for alle løsninger unntatt Alternativ 0 (som ikke er holdbar) og Alternativ 0+ (som vi ikke anbefaler).

Denne programvareoppdateringen er i prinsippet en svært enkel prosedyre, og vil ha en tilnærmet lik kostnad *uansett hvilken ny identifikator man til slutt velger*. I praksis kan mange aktører likevel føle og gi uttrykk for et ubehag ved å gjennomføre disse oppdateringene, fordi programvaren som utfører kontrollen kan være svært gammel og dårlig dokumentert slik at endringer kan gi uforutsette bivirkninger.

3.10 Teknologi, datasikkerhet og personvern

I det halve århundret som har gått siden dagens fødselsnummerordning ble innført, har samfunnet gått gjennom en utvikling, både teknologisk og med hensyn på befolkningsvekst, som det var svært vanskelig å forutse på forhånd. På samme måte som det er vanskelig å gi pålitelige prognoser for framtidig befolkningsvekst (som diskutert i avsnitt 3.2), er det også vanskelig å spå hvordan teknologien vil påvirke rutiner for personregistrering.

Generelt kan man si at dersom det er ønskelig (av hensyn til datasikkerhet og personvern) å knytte for eksempel biometriske data til en personidentifikator, eller dersom man ønsker å ta i bruk identitetsbasert kryptografi, kan det være ønskelig å utvide rommet av tilgjengelige identifikatorer betydelig. Det kan bety at man ønsker en identifikator av større lengde enn det som er aktuelt nå, som benytter et stort alfabet. Varianter av Alternativene 2 eller 4 kan ivareta slike ønsker.

4) Svar på oppgaver

I dette avsnitte vil vi svare på de konkrete spørsmål som definerer oppdraget vårt.

4.1 Vurdering av kapasitet i eksisterende identifikator

Verifisering og korreksjon av de vurderinger som er gjort i forespørselens kapittel 2 angående egenskaper og kapasitet for dagens fødselsnummer.

Tabell 1 og Tabell 2 nedenfor viser sammenhengen mellom det første siffer i individnummer og fødselsår.

Første siffer ↓	To siste siffer i årstall		
	00-39	40-54	55-99
0	2000-2039	1940-1954	1955-1999
1	1900-1939	1940-1954	1955-1999
2			
3			
4			
5	2000-2039	2040-2054	1855-1899
6			Ikke i bruk
7			
8		1940-1954 (Reserve)	1955-1999 (Reserve)
9		2040-2054 (Hvis ikke brukt)	Ikke i bruk

Tabell 1 Koding av århundre. For hvert første siffer er det i gjennomsnitt 82,6 gyldige løpenumre pr fødselsdato.

Første siffer i individnr.	Fødselsår	Født i perioden	Nominelt antall individnr.	Gjennomsnittlig antall gyldige individnr. per dag
5, 6, 7	> 54	1855-1899	749-500	206,6
1, 2, 3, 4		1900-1999	499-100	330,6
0	> 39	1940-1999	099-000	82,6
9 (Reserve)	> 39	1940-1999	999-900	82,6
5, 6, 7, 8	< 55	2000-2054	899-500	330,6
9	< 40	2000-2039	999-900	82,6
0	< 40	2000-2039	099-000	82,6
9 (Reserve, hvis 9 ikke brukes 1940-1999)	> 39	2040-2054	999-900	82,6

Tabell 2 Sammenhengen mellom det første siffer i individnummer og fødselsår

Ikke alle individnumre kan benyttes da enkelte ikke godkjennes av moduluskontrollen som er bygd inn for å avsløre eventuelle feil i fødselsnummeret. Av en serie på 100 numre forkastes ca. 17 individnumre ($10/11 \cdot 10/11 = 100/121 = 82,6\%$). Derfor er det bare 82,6 tilgjengelige individnumre i en serie på 100. Se forøvrig Appendiks 2.

Hvis reserven (999-900) for perioden 1940-1954(1999) ikke brukes, vil systemet for fødselsnummer ha en estimert kapasitet på 413 ($330,6 + 82,6$) individnumre per dato for årene 1940-2054. I perioden 2000-2039 vil kapasiteten øke med 83 til 496 individnumre fordi serien 099-000 kan brukes. Fra og med 2040 vil det ikke være mulig å skille nye fødselsnumre fra perioden 1940-1999 for individnumre i serien 099-000. Kapasiteten faller dermed tilbake til 413 individnumre per dato etter 2039.

Hvis derimot reserveserien tas i bruk for tidsperioden 1940-1954(1999), dvs. tildeles personer som kommer til Norge gjennom innvandring etc., vil kapasiteten fra og med 2040 kunne bli redusert ned mot 330 individnumre per dato.

Nedenfor er det gjort en oppstilling av tilgjengelig kapasitet av individnumre for de ulike tidsperioder:

	Første siffer i individualsiffer										Kapasitet per dag
	0	1	2	3	4	5	6	7	8	9	
1855-1899	-	-	-	-	-	83	83	41	-	-	207
1900-1939	-	83	83	83	83	-	-	-	-	-	331
1940-1999 (eks. reserve)	83	83	83	83	83	-	-	-	-	-	413
1940-1999 (inkl. reserven)	83	83	83	83	83	-	-	-	-	83	496
2000-2039	83	-	-	-	-	83	83	83	83	83	496
2040-2054 (eks. reserve)	-	-	-	-	-	83	83	83	83	-	331
2040-2054 (inkl. reserve)	-	-	-	-	-	83	83	83	83	83	413

Tabell 3 Antall tilgjengelige individnumre per dato for tidsperioder

I Statistisk sentralbyrås prognoser, diskutert i Appendiks 5, vil det ved **middels** nasjonal vekst være dager der antall fødte inkludert innvandring overstiger kapasiteten for året 2040 dersom reserveserien er benyttet tidligere. Se Figur A5.2 i Appendiks 5. Ved disse prognosene er det grunn til å tro (men ikke sikkert) at kapasitet inkludert reserveserie kan være såvidt tilstrekkelig, også fra 2040. Uten reserveserien vil kapasiteten være for liten fra 2040.

Per desember 2011 er reserveserien ennå ikke blitt benyttet.

Statistisk sentralbyrås prognoser viser at ved **høy** nasjonal vekst vil antall fødte inkludert innvandring overstige kapasiteten allerede fra 2014, selv om reserveserien skulle være tilgjengelig.

Hvis det i perioden fram til 2040 skulle skje utvidelse av registerpopulasjonen, dvs. endrede forutsetninger gjennom at nye personkretser blir registrert i DSF, vil det bli nødvendig å erstatte dagens fødselsnummerordning fra det tidspunkt utvidelsen av registerpopulasjonen finner sted, avhengig av hvor stor utvidelsen blir.

4.2 Numeriske tegn kontra alfanumeriske tegn

Hvilke konsekvenser (fordeler og ulemper) gir numeriske tegn kontra alfanumeriske tegn. Vil valg av tegntype ha betydning for moduluskontroll, se punkt 3 nedenfor.

- Alfnumeriske tegn utvider alfabetet og dermed kapasiteten.
- Hvert tegn får en numerisk verdi (f. eks. 0-9, A=10, B=11, C=12 etc.) som brukes direkte (og problemfritt) i modulus kontrollen.
- Moduluskontroll bruker et primtall q som er minst like stor som antall symboler i alfabetet.
 - For eksempel: med 13 alfanumeriske tegn 0-9, A=10, B=11, C=12 kan vi bruke modulus kontroll med $q=13$.
- Fordeler med numeriske (desimale) tegn
 - Trenger bare numerisk tastatur for å registrere
 - **Tydelig**; begrenset risiko for feilregistrering
 - Men ikke perfekt. Se avsnitt 3.5.
 - (For noen) lettere å huske?
- Ulemper med numeriske tegn
 - Mye mindre kapasitet (men det er uklart hva som menes med «alfanumerisk». Se nedenfor.)

- Desimalt alfabet: Alfabetstørrelse er ikke primtallspotens, taper mange mulige sekvensnumre

Oppsummering: Det er i dag ikke signifikante fordeler ved desimale systemer i forhold til systemer med større alfabet.

Hva menes med «alfanumerisk»?

En vanlig definisjon vil være at et fullt alfanumerisk alfabet har symbolene 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z.

Her har vi tatt vekk Æ, Ø, Å fordi det er veldig gode grunner til å unngå disse, blant annet at utenlandske tastaturer mangler disse symbolene, og at alle disse tre symbolene lett kan forveksles med andre både ved skriftlig og muntlig overføring (Æ/E, Ø/O, Å/A).

I det fulle alfanumeriske alfabetet er det 36 symboler. Som nevnt i Appendix 1 er det en ulempe at alfabetstørrelsen ikke er et primtall. Dette kan vi løse på følgende måter:

- Alfabetstørrelse bør være, for å unngå tap av lovlig identifikatorer,
 - et primtall (11, 13, 17, 19, 23, 29, 31, ...)
 - eller en primtallspotens ($16=2^4$, $25=5^2$, $27=3^3$, $32=2^5$, ...)
 - helst et primtall (enklere å beskrive algoritmen!)
- Kan velge en undermengde, med et primtalls antall elementer
 - Ved å velge denne undermengden (alfabetet) på en passende måte kan vi også redusere sannsynlighet for feil ved registrering, ved å unngå å ha med par av symboler som kan forveksles verbalt eller skriftlig, eller som er naboer på et vanlig QWERTY-tastatur.
- Eksempler (ikke optimalisert!):
 - 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, C, D, E, F, H, J, K, L, M, N, P, R, T, U, V, W, X, Y. (29 tegn)
 - 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, D, E, F, H, K, M, N, P, V, W, X, Y. (23 tegn)

4.3 Feilkontroll

Hvor mange kontrollsifre anbefales benyttet og med hvilken moduluskontroll. Hvis antall kontrollsifre og moduluskontroll anbefales forskjellig for ulike personidentifikatorer, skal analysen beskrive disse forskjeller med anbefaling for grupper av alternative personidentifikatorer, eventuelt for den enkelte personidentifikator.

Generelt kan vi si følgende:

- Ved bruk av **ett kontrollsiffer** (symbol) kan vi, ved riktig valg av metode, oppdage alle enkle feil og alle ombyttinger av to posisjoner.
- Ved bruk av **to kontrollsiffer** (symbol) kan vi, ved riktig valg av metode, oppdage alle doble feil, og alle ombyttinger av to posisjoner. *Denne egenskapen er ikke oppfylt i dagens fødselsnummer.*
I tillegg er noe kapasitet tilgjengelig til å oppdage ytterligere aktuelle feil som kan kreve en mer detaljert analyse, avhengig av konkrete problem og prioriteringer.
- Muligheter for feilkorreksjon: dersom en identifikator er feil registrert kan et dataprogram produsere en kort liste av sannsynlig korrekte fødselsnumre.

- Både feildeteksjons- og feilkorreksjonsegenskaper forbedres betydelig med økende antall kontrollcifre. Vi anser det imidlertid ikke naturlig for denne anvendelsen å anbefale mer enn to kontrollcifre.
- Dersom man har et eller flere kontrollsymbol, gir det mulighet for fremtidig utvidelse av kapasitet ved å frigjøre ett kontrollsymbol. Dette gir en kostnad i form av redusert pålitelighet. Selv om man har flere kontrollsymbol igjen etter frigjøringen, er det ikke sikkert at disse er designet for å stå alene; dette gjelder blant annet dagens fødselsnummerordning, der siste kontrollsymbol ikke er et optimalt valg hvis systemet skulle blitt utviklet ”fra scratch” i dag.

Appendiks 1 gir en mer utfyllende innføring i bakgrunn for feildetekterende koder for personidentifikatorer. I Appendiks 2 beskriver vi hvordan denne teorien er anvendt i dagens fødselsnummer.

4.4 Oversikt over alternativer

Beregning av fremtidig kapasitet for de alternative nye personidentifikatorene som er beskrevet i forespørselens kapittel 3. Med kapasitet menes antall mulige unike identifikasjoner for den enkelte identifikator per dato. Her må det tas hensyn til eventuell bortfall pga. avvisninger i forbindelse med moduluskontroll etc.

Tabell 4 gir en oversikt over egenskaper ved de alternativene vi har studert, målt i forhold til kriteriene som er beskrevet i Avsnitt 3). Appendiks 3 inneholder en detaljert forklaring av vurderingene av hvert alternativ.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.0 DDMMÅÅIIKK	Ja	--	+		des	++	-	+	+	
Alt. 0+ DDMMÅÅIIKK	Tja	-	+		des	++	-	--	+	
Alt. 0X DDMMÅÅXXXXKK	Tja	-	+		11	+	-	+	0	
Alt.1 DDMMÅÅIIKK	Ja	++	0+		des		0	+	-	
Alt. 2 DDMMÅÅAAKK	Ja	++	++		q	++	+	+	-	+
Alt. 3 Kundenummer IIIIIIIIKK	Nei	++	+	-	des	0	-	+	-	
Alt. 4 AAAAAAK(K)	Nei	++	++	+	q	0	++	+	-	++
Alt.5 Skala måned DDMMÅÅIIKK	Tja	0	+		des		+	+	-	
Alt. 6 14 sifre DDMMÅÅÅÅIIKK	Ja	++	0	--	des			+	-	
Alt. 7 Svensk løsning IIMMÅÅIIKK	Tja	+	+		des	-	+	+	-	

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt. 8 DDMMÅÅAIIKK	Ja	--	-		<i>q</i>	-	-	+	-	-
Tegnforklaring: ++ Veldig bra, + Bra, 0: Nøytralt, - : Ulempe, -- Uakseptabelt, Blank: Ikke vurdert										

Tabell 4 – Sammendrag av egenskaper for forskjellige identifikatorsystem.

Ranking

Informasjonsløs identifikator

I realiteten er det bare to informasjonsløse alternativ blant de vi har vurdert: Alternativ 3 (desimale symboler) og Alternativ 4 (alfanumeriske symboler). Dersom man skal skifte til en informasjonsløs identifikator av *politiske* årsaker, bør alle i populasjonen få nye identifikatorer. Da vil det være en ulempe snarere enn en fordel om den nye identifikatoren har et format som ligner på den gamle. Alternativ 4 er betydelig kortere og lettere å huske enn alternativ 3, og dersom man velger to kontrollsiffer vil Alternativ 4 også ha betydelig bedre feilkontroll. Alternativ 3 har større kapasitet (760 millioner), mens Alternativ 4 (6 løpenummersymboler, $q=23$) har kapasitet på 148 millioner. Men begge er stor nok, og Alternativ 4 kan med enkle grep utvides til å bli *mye* større og fremdeles være enklere enn Alternativ 3.

Alternativ 4 (med lengde 8, og to kontrollsymbol) er mer robust mot feil enn Alternativ 3 (lengde 11, to kontrollsymbol), som vist i Figur 1.

Vi anbefaler Alternativ 4 med to kontrollsymbol som informasjonsløs identifikator.

Identifikator med informasjonsinnhold

Blant de alternativene med dato og annen informasjon innbakt vil vi eliminere Alt. 0, Alt. 0X, Alt. 8 av rene kapasitetshensyn, Alt. 0+ fordi det mangler entydighet og av kapasitetshensyn, og Alt. 6 fordi identifikatoren blir ekstra lang uten at dette gir en tilstrekkelig stor kapasitetsgevinst.

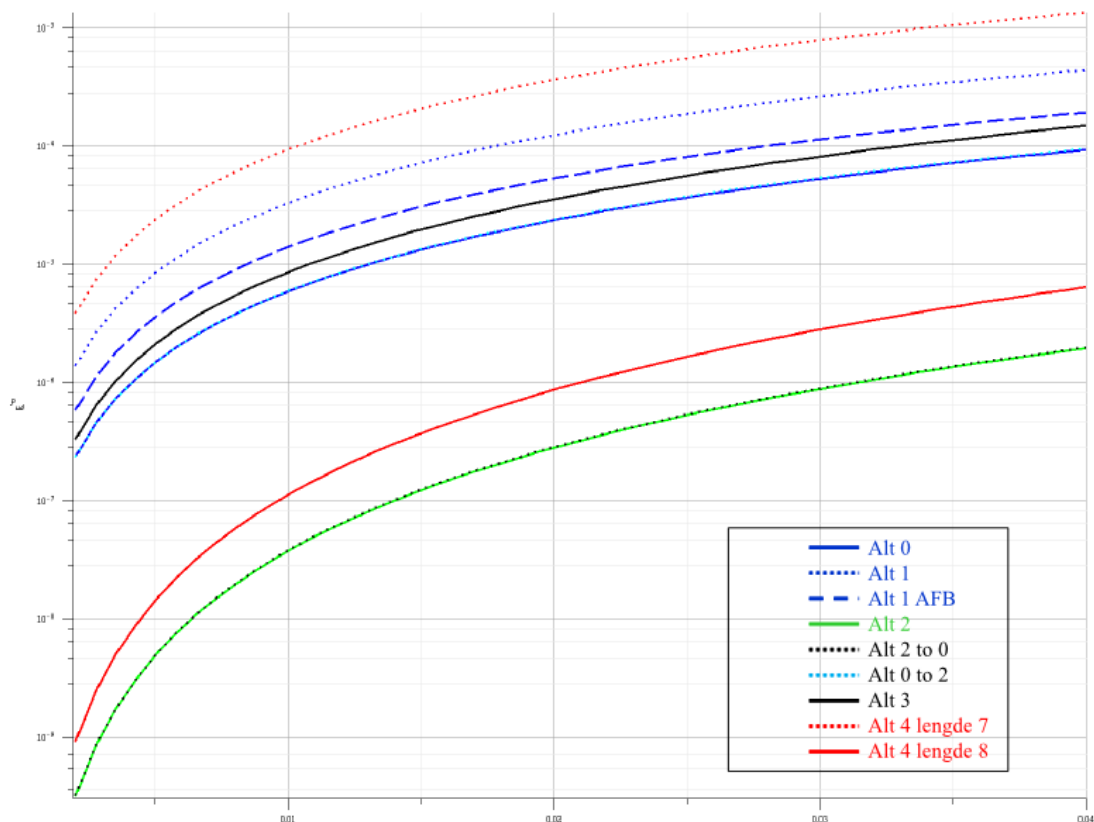
De gjenværende alternativene er da Alt. 1, Alt. 2, Alt 5, Alt. 7. Alt. 5 og Alt. 7 gir sannsynligvis tilstrekkelig kapasitet i forhold til kravet om varighet til 2150, men kapasiteten i disse alternativene er betydelig lavere enn i de to andre. Alt. 5 og Alt. 7 har også en "grumset" dato innbakt.

Vi anser derfor Alternativene 1 og 2 (variant 2, med $q=23$, som beskrevet i Appendiks 3) som de teknisk beste, og de har tilnærmet samme kapasitet i utgangspunktet, og denne kapasiteten varer i mange hundre år med dagens prognoser for befolkningsvekst. Alternativ 2 har en bedre feilkontroll og har et større potensial for enkle fremtidige utvidelser. Alternativ 1 ligger nærmere opp til dagens løsning med hensyn til alfabet og feilkontroll.

Figur 1 viser sannsynlighet for uoppdaget feil ved Alternativ 1 kontra Alternativ 2. Som man ser er Alternativ 2 mer robust, men fornuftige tildelingsprosedyrer for identifikatorer i Alternativ 1

kan gjøre at feildeteksjonsegenskapene i Alternativ 1 ligger nær de man er vant til med dagens fødselsnummer.

Ut fra rent tekniske betraktninger anser vi Alternativ 2 som en vinner foran Alternativ 1. Men det kan være attraktivt at Alternativ 1 ligner mye på dagens fødselsnummer, og i en helhetsvurdering kan Alternativ 1 derfor være å foretrekke.



Figur 1 : Figuren viser sannsynligheten for uoppdaget feil, P_{ud} , som funksjon av p , der $p/(q-1)$ er sannsynligheten for at et symbol blir feiltolket som et av de andre $q-1$ lovlige tegnene i alfabetet og $(1-p)$ er sannsynligheten for feilfri tolkning av et enkelt symbol (dette er en såkalt q -nær kanal), ved bruk av henholdsvis Alternativ 4 med lengde 8, Alternativ 4 med lengde 7 (ett kontrollsymbol), Alternativ 3, Alternativ 0 (dagens system), Alternativ 1, Alternativ 2 (versjon 2) frittstående, Alternativ 2-identifikator samtidig med Alternativ 0 og Alternativ 0 samtidig med Alternativ 2. I tillegg viser vi sannsynligheten P_{AFB} for Alternativ 1. Se Appendiks 3 for forklaring.¹

4.5 Praktisk gjennomføring

Forklar hvordan dette praktisk kan gjennomføres.

Ved bruk av en informasjonsløs identifikator forutsetter vi i utgangspunktet at hele populasjonen får tildelt en ny identifikator fra innføringstidspunktet, fordi vi antar at et slikt skifte i første rekke ville være motivert ut fra bekymringer om personvern. Dersom man likevel ønsker å benytte dagens fødselsnummer parallelt med en av de foreslåtte informasjonsløse identifikatorene, er det likevel ikke noe til hinder for dette.

¹ Figur 1 er beregnet på grunnlag av såkalte vektfordelinger for de forskjellige identifikatorene. Vi beskriver ikke i dette dokumentet detaljene som ligger til grunn for Figur 1.

Ved bruk av en identifikator som inneholder informasjon er det et ønske at de som har fått et fødselsnummer etter gammel ordning får beholde dette (se avsnitt 3.6.) De forskjellige alternativene varierer med hensyn til hvor lett dette er å få til, og hvordan det kan gjøres. Vi henviser til Appendiks 3 for en diskusjon av de enkelte alternativ.

4.6 Andre alternativ

Det er i forespørselens kapittel 3 beskrevet åtte alternative nye personidentifikatorer. Hvis utreder ser andre mulige personidentifikatorer som anbefales utredet, bes det om at disse defineres og kort beskrives med hensyn til egenskaper etc.

Alternativene som er studert i Appendiks 3 og oppsummert i Avsnitt 4.4, er i noen tilfeller modifisert i forhold til de opprinnelige forslag.

Appendiks 1: Innføring i feildetekterende koder

Feilkorrigerende (eller feildetekterende) koder brukes til å oppdage (eller korrigere) feil ved overføring eller lagring av data. Idéen er å tilføre redundans i form av kontrollcifre til dataene slik at dersom feil oppstår kan vi oppdage og eventuelt korrigere feilene.

Generelt kreves det flere kontrollcifre for å korrigere en feil enn å bare oppdage den. I dagens personnummersystem brukes to kontrollcifre som oppdager flere typer av feil. Dette inkluderer enhver feil i et enkelt siffer og alle ombyttinger av data i naboposisjoner og en del andre typer feil som ombytting av måned og dato.

Dess flere kontrollcifre en bruker dess flere feil kan oppdages. Ulempen er at kontrollcifre øker redundansen og ikke øker informasjonskapasiteten.

I de fleste nasjonale system brukes 0, 1 eller 2 kontrollcifre. Vi vil derfor beskrive mer detaljert hvorledes vi kan bruke 1 og 2 kontrollcifre mest effektivt til feilkontroll. Tre eller flere cifre anses som unødvendig for denne anvendelsen. Praktisk er det ingen problemer å implementere flere kontrollcifre.

Vi antar at personidentifikatoren $p_1, p_2, \dots, p_n$ har symboler som alle er hentet fra et alfabet $\mathcal{A}$ med $s(\mathcal{A})$ symboler. Disse symbolene tilordnes numeriske verdier i mengden $\{0, 1, \dots, s(\mathcal{A})-1\}$.

Prinsippet for feilkontroll

Feilkontrollcifrene i dagens personnummer er basert på modulus q kontroll. Prinsippet er meget enkelt: Anta at $s(\mathcal{A}) = q$, der q et primtall. Anta videre vi har et nummer med $p_1, p_2, p_3, \dots, p_{n-1}$ og at vi skal lage et gyldig kontrollsiffer $K = p_n$ slik at vi får et gyldig nummer $p_1, p_2, p_3, \dots, p_n$. Vi velger vektor $w_1 w_2 w_3 \dots w_n$ i mengden $\{1, 2, \dots, q-1\}$, det vil si at alle vektorer er ulike 0. Vi beregner kontrollsifferet K ved å løse ligningen nedenfor slik at

$$S = w_1 \cdot p_1 + w_2 \cdot p_2 + w_3 \cdot p_3 + \dots + w_{n-1} \cdot p_{n-1} + w_n \cdot K$$

er **delelig med** q . Når q er et **primtall** har dette alltid en løsning for K i mengden $\{0, 1, \dots, q-1\}$.

Redundans innføres siden alle *gyldige* personidentifikatorer $p_1, p_2, p_3, \dots, p_n$ (der $K=p_n$) har en verdi av S som er delelig med q og derfor er bare $1/q$ av alle kombinasjoner av n symboler gyldige. Prinsippet for feilkontroll er at alle feil som transformerer en gyldig identifikator til en tegnstreng der denne kontrollsummen *ikke* er delelig med q , vil oppdages.

Vi lar $S \equiv i \pmod{q}$ bety at S gir i til rest ved divisjon med q . Spesielt vil $S \equiv 0 \pmod{q}$ bety at S er delelig med q .

Bruk av ett kontrollsiffer

Enkle feil oppdages når ingen av vektene er delelige med q . Alle enkle feil i personidentifikatoren kan oppdages når q er et primtall og ingen av vektene er 0. Dette skyldes at en gyldig personidentifikator $p_1, p_2, p_3, \dots, p_n$ (der $K=p_n$) oppfyller

$$S = w_1 \cdot p_1 + w_2 \cdot p_2 + w_3 \cdot p_3 + \dots + w_{n-1} \cdot p_{n-1} + w_n \cdot K \equiv 0 \pmod{q}$$

Hva skjer om en enkel feil opptrer i posisjon nummer t , for eksempel at p_t endres til $p_t + e$, der e er ulik 0? Dette gir et nytt nummer $p_1, p_2, p_3, \dots, p_t + e, \dots, p_n$ (der $K=p_n$). Ved å beregne sjekksummen S' for dette nummeret får vi, siden $S \equiv 0 \pmod{q}$ at

$$\begin{aligned} S' &= w_1 \cdot p_1 + w_2 \cdot p_2 + \dots + w_t \cdot (p_t + e) + \dots + w_n \cdot p_n \\ &= w_1 \cdot p_1 + w_2 \cdot p_2 + \dots + w_t \cdot p_t + \dots + w_n \cdot p_n + w_t \cdot e \\ &= S + w_t \cdot e \\ &\equiv w_t \cdot e \pmod{q}. \end{aligned}$$

En enkel feil e i posisjon nummer t endrer kontrollsjekken med $w_t \cdot e$ som **ikke kan være delelig med primtallet q** siden w_t og e begge er mellom 1 og $q-1$ og ingen av dem er delelig med q . (Denne egenskapen er ikke oppfylt dersom q ikke hadde vært et primtall.)

Alle ombyttinger oppdages når alle vektene er forskjellige: Anta at vi har en gyldig personidentifikator $p_1, p_2, p_3, \dots, p_n$ (med $K=p_n$) med symboler fra et alfabet med q symboler $\{0, 1, \dots, q-1\}$ der kontrollsymbolet er valgt slik at

$$S = w_1 \cdot p_1 + w_2 \cdot p_2 + w_3 \cdot p_3 + \dots + w_n \cdot p_n \equiv 0 \pmod{q},$$

Det vil si at S er delelig med q . Dersom vi bytter om symbolene i to posisjoner, i og j , og regner ut sjekksummen S' for det ombyttete nummer, så vil kontrollsjekken gi

$$S' = S + (p_i - p_j) \cdot (w_j - w_i),$$

som heller ikke kan være delelig med q når $p_i \neq p_j$ og vektene $w_i \neq w_j$ for alle i og j .

Siden vi vil studere personidentifikatorer fra et vilkårlig alfabet er det viktig å merke seg følgende. La personidentifikator inneholde elementer fra et alfabet $\mathcal{A}$ med $s(\mathcal{A})$ symboler. Vi tilordner numeriske verdier $0, 1, \dots, s(\mathcal{A})-1$ til de $s(\mathcal{A})$ symbolene i alfabetet.

1. Dersom $s(\mathcal{A})$ er et primtall og alle vekter er ulik 0 så oppdages alle enkle feil i en posisjon.
2. Dersom $s(\mathcal{A})$ er et primtall og alle vekter er innbyrdes forskjellige så vil alle ombyttinger av to posisjoner oppdages. Spesielt må den totale lengden av personidentifikatoren være mindre enn primtallet q som brukes i moduluskontrollen.

3. Fra 1) og 2) følger at vi kan oppdage alle enkle feil og alle ombyttinger ved å velge alle vektorer innbyrdes ulike (mod q) og ulike 0. Spesielt må den totale lengden av personidentifikatoren være mindre enn primtallet q som brukes i moduluskontrollen.²

Konklusjon om optimal bruk av ett kontrollsiffer: Dersom vi velger alfanumeriske tegn i personidentifikatoren fra et alfabet med høyst q symboler, der q er et primtall, kan vi ved hjelp av **ett kontroll symbol** oppdage alle enkle feil og **samtidig oppdage alle ombyttinger** av to forskjellige posisjoner. Dette krever at alle vektorer er ulike 0 og innbyrdes forskjellige, det vil si at den totale lengden er mindre enn q .³

Det ideelle når man bruker ett kontrollsymbol er å ha kontroll modulo q , der q er primtall som er større enn lengden siden dette oppdager alle feilene ovenfor. (Det er også mulig å la q være en primtallspotens, men da er "modulus" kontrollen mer komplisert).

Bruk av to kontrollsymboler: Ved hjelp av **to** kontrollsymboler kan vi oppdage alle feil i to posisjoner, dersom vi koder optimalt. La personidentifikatoren bestå av symbolene $p_1 p_2 p_3 \dots p_n$ der $K_1 = p_{n-1}$ og $K_2 = p_n$ er de to kontrollsymbolene. La $w_1, w_2, \dots, w_n$ være de tilsvarende vektene i beregningene av den første kontrollsymbolet K_1 og la $v_1, v_2, \dots, v_n$ være vektene i beregningene av det andre kontrollsymbolet K_2 . (Vi kan anta at $w_n = 0$.) Da velges K_1 og K_2 slik at sjekksummene nedenfor begge er delelig med q , det vil si slik at

$$S_1 = w_1 \cdot p_1 + w_2 \cdot p_2 + w_3 \cdot p_3 + \dots + w_{n-2} \cdot p_{n-2} + w_{n-1} \cdot K_1 = 0 \pmod{q}$$

og

$$S_2 = v_1 \cdot p_1 + v_2 \cdot p_2 + v_3 \cdot p_3 + \dots + v_{n-2} \cdot p_{n-2} + v_{n-1} \cdot K_1 + v_n \cdot K_2 = 0 \pmod{q}.$$

For at denne metoden skal oppdage alle doble feil kan vi vise ved enkle kodeteori-metoder at da må alle v_i/w_i være forskjellige (mod q) for $i=1,2,\dots,n$.⁴

Bruk av mer enn to kontrollsymboler: Dersom vi ønsker å oppdage alle kombinasjoner av t feil, der t er større enn 2, er den beste generelle metoden å benytte *Reed-Solomon*-koder. Ved å bruke t kontrollsymboler kan vi oppdage alle feil i t posisjoner. Dette oppnås enkelt av Reed-Solomon-koder som er enkle å implementere (for eksperter) og som kan brukes når antall symboler er en primtallspotens (det vil si, q er på formen p^k for et primtall p). Vi vil likevel foretrekke å bruke tilfellet når alfabetet er et primtall, siden reglene for kontrollberegninger er noe enklere å implementere (og å forklare!).

Hvor mange kontrollsymboler bør benyttes? Dette avhenger av hvor sterk feildeteksjon det er behov for og av valget av feildeteksjonskode. Populært kan vi si at **ett kontrollsiffer** oppdager alle feil i en enkel posisjon og i tillegg alle ombyttinger av to symboler. **To kontrollsifre** oppdager alle kombinasjoner av to feil i personidentifikator. I tillegg er det muligheter for å oppdage andre viktige feil, som for eksempel ombytting av måned og dato. Det vil generelt være nødvendig med en mer detaljert skreddersøm av koden for å korrigere slike tilleggsfeil.

² Den observante leser vil registrere at dette *ikke* gjelder dagens fødselsnummersystem, der både lengde og grunntall q er lik 11. Men dagens fødselsnummer vil ikke oppdage alle feil i to posisjoner, og heller ikke alle ombyttinger.

³ I noen tilfeller kan vi øke lengden til å være lik q eller $q+1$. For å forenkle fremstillingen lar vi være å gå i detalj.

⁴ Vi definerer at $v_n/w_n = v_n/0 = \infty$.

Appendiks 2: Beregning av kontrollsiffer i dagens fødselsnummer

Dagens fødselsnummer er som nevnt på formen

$$D_1 D_2 M_1 M_2 \mathring{A}_1 \mathring{A}_2 I_1 I_2 I_3 K_1 K_2$$

og kontrollsifferne beregnes essensielt som beskrevet i Appendiks 1. Forskjellen er at kontrollsiffer beregnes modulo 11, mens alle symboler er desimale, altså i alfabetet $A_{10} = \{0,1,2,3,4,5,6,7,8,9\}$.

K_1 blir bestemt ved at summen

$$S_1 = 3 \cdot D_1 + 7 \cdot D_2 + 6 \cdot M_1 + 1 \cdot M_2 + 8 \cdot \mathring{A}_1 + 9 \cdot \mathring{A}_2 + 4 \cdot I_1 + 5 \cdot I_2 + 2 \cdot I_3 + 1 \cdot K_1$$

tvinges til å være lik 0 (mod 11). Dette gir en ligning som for bestemte verdier av $D_1 D_2 M_1 M_2 \mathring{A}_1 \mathring{A}_2 I_1 I_2 I_3$ gir en entydig løsning for K_1 . Av og til (det vil si i 1/11 av tilfellene) gir verdien $K_1 = 10$, i disse tilfellene blir løpenummeret $I_1 I_2 I_3$ forkastet. I notasjonen fra Appendiks 1 har vi altså

$$w_1 = 3, w_2 = 7, w_3 = 6, w_4 = 1, w_5 = 8, w_6 = 9, w_7 = 4, w_8 = 5, w_9 = 2, \text{ og } w_{10} = 1.$$

På samme måte blir K_2 blir bestemt ved at summen

$$S_2 = 5 \cdot D_1 + 4 \cdot D_2 + 3 \cdot M_1 + 2 \cdot M_2 + 7 \cdot \mathring{A}_1 + 6 \cdot \mathring{A}_2 + 5 \cdot I_1 + 4 \cdot I_2 + 3 \cdot I_3 + 2 \cdot K_1 + 1 \cdot K_2$$

settes lik 0 (mod 11), og at løsningen løses med hensyn på K_2 .

Som nevnt før betyr dette at for hver DDMMÅÅ er det bare 826 gyldige valg av identifikatorverdier. Koden detekterer alle enkle feil, men *ikke alle* doble feil.

Appendiks 3: Oversikt over mulige identifikatorer

I arbeidet med konseptvalgutredningen har Skattedirektoratet/prosjektet gjennom en grovsiling kommet fram til noen foreløpige alternative nye personidentifikatorer som er tenkt å fange opp flere scenarier med hensyn til behov for kapasitet og egenskaper. Skattedirektoratet har satt som krav at en ny personidentifikator skal ha tilstrekkelig kapasitet for anvendelse minst fram til år 2150. Det er også knyttet usikkerhet til hvor mange kontrollsifre som er tilstrekkelig for en ny personidentifikator.

Alternative personidentifikatorer kan bl.a. grupperes etter følgende hoveddimensjoner:

- Tegntype – kun numeriske tegn eller bruk av alfanumeriske tegn
- Informasjonsbærende eller informasjonsløs. Dagens fødselsnummer betegnes som informasjonsbærende og gir informasjon om både fødselsdato og kjønn. Informasjonsløs innebærer en tilfeldig rekke av tall eller bokstaver uten konkret informasjon om personen.
- Lengden, dvs. antall posisjoner på identifikatoren – dagens fødselsnummer har 11 posisjoner (11 sifre).
- Antall kontrollsifre og tilhørende moduluskontroll – dagens fødselsnummer har to kontrollsifre som beregnes etter modulus 11.

I tabellen nedenfor har vi tatt utgangspunkt i de to dimensjonene tegntype (numeriske eller alfanumeriske tegn) og informasjon (informasjonsbærende/informasjonsløs) og gruppert deretter identifikatorene i fire hovedgrupper. De andre dimensjonene fremgår av forklaringen for den enkelte identifikator. Totalt er det åtte personidentifikatorer i tillegg til dagens fødselsnummer:

Informasjon Tegntype	Informasjonsbærende	Informasjonsløs eller med begrenset informasjon
Numeriske tegn	0. Dagens fødselsnummer (11): DDMMÅÅIIKK 1. Ett kontrollsiffer (11): DDMMÅÅIIK 2. Skala for måned utvides (11): DDMMÅÅIIKK(IK) 3. 14 siffer: DDMMÅÅÅÅIIKK (K)	4. Dag inngår i individnr., svensk løsning (11): IIMMÅÅIIKK 5. Kundenummer (11): IIIIIIIIKK
Alfanumeriske tegn	6. Frigjort K bokstav og angir århundre (11): DDMMÅÅIIBK 7. Individposisjoner alfanum., angir også århundre (11): DDMMÅÅIIKK (I=A)	8. Informasjonsløs og alfanumerisk (8): AAAAAAKK (K)

Tabell 5: Alternative personidentifikatorer

A=alfanumerisk, B=bokstav, D=dag, M=måned, Å=år, I=individnr., K=kontrollsiffer

Nedenfor er alternativene kort forklart. **På grunn av kompatibilitet med fremtidige dokument i saksbehandlingen er alternativene i tabellen ovenfor renummerert.** Spesifikt er Alternativ 2, i tabellen ovenfor, renummerert til Alt. 5, Alt. 3 er renummerert til Alt. 6, Alt. 4 er renummerert til Alt. 7, Alt. 5 er renummerert til Alt. 3, Alt. 6 er renummerert til Alt. 2, Alt. 7 er renummerert til Alt. 8, og Alt. 8 er renummerert til Alt. 4.

Informasjonsbærende med numeriske tegn:

Alt 0. - Nullalternativet. Dagens fødselsnummer - 11 posisjoner, 2 kontrollsifre

En videreføring av dagens løsning kalles nullalternativet.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.0 DDMMÅÅIIKK	Ja	--	+		des	++	-	+	+	

Som diskutert i Avsnitt 4.1 er det fare for at kapasiteten i dagens fødselsnummersystem er oppbrukt innen få år.

Sannsynlighet for uoppdaget feil er vist i Figur 1 (heltrukken blå kurve.)

Alt 0+: Dagens fødselsnummer - 11 posisjoner, 2 kontrollcifre

Det kan vurderes å fjerne begrensning om å benytte brukte nummerserier, f.eks. for døde personer født før 1900. Dette alternativet kan kalles 0+ (Nullalternativet+).

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt. 0+ DDMMÅÅIIKK	Tja	-	+		des	++	-	--	+	

Det finnes flere varianter av denne løsningen.

- *Vi antar at man kan bruke om igjen løpenummer for døde personer født før 1900.*
I praksis har dette ingen betydning før perioden 2055-2099 (se Tabell 1). Da kan man ha tilgjengelig det samme antall identifikator som for perioden 2040-2054, det vil si 413 pr. dag.
- *Vi antar at man kan bruke om igjen løpenummer for personer som har vært døde i x år, uansett fødselsår.*
Med forventet lengre levetid antar vi at en stor del av hvert årskull vil leve mer enn 100- x år. Dersom vi forholder oss til en gjennomsnittlig levealder på 80 år, vil medianen antagelig være over 80 år. Det vil si at for $x=20$, for eksempel, vil mindre enn halvparten av forrige århundres fødselsnummer være tilgjengelig (for å si det veldig brutalt) for tildeling etter dette prinsippet. Hvis vi velger $x=20$ vil derfor dagskapasiteten likevel være mindre enn 620. Uansett valg av verdi på x vil dagskapasiteten ikke kunne overstige 826.

Den åpenbare ulempen med dette alternativet er at **entydigheten forsvinner**. En annen ulempe er at det **ikke er mulig lenger å kode for århundre**. Løsningen vil uansett være marginal på sikt.

Alt. 1. Dagens fødselsnummer med ett kontrollsiffer - 11 posisjoner

Frigjort K går inn i individnummer. Dette er Grunndatarapportens subsidiære forslag.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.1 DDMMÅÅIIKK	Ja	++	0+		des		0	+	-	

Vi forutsetter samme feilkontroll på det gjenværende kontrollsymbolet (K_2) som i dag. Dette alternativet er altså på formen DDMMÅÅ $I_1I_2I_3I_4K_2$, mens de gamle/eksisterende identifikatorene er på formen DDMMÅÅ $I_1I_2I_3K_1K_2$.

Kapasiteten for hver dag DDMMÅÅ vil være 8264.

Koding av århundre:

- Koding av århundre gjennom bruk av første siffer I_1 (Tabell 1, Tabell 2) er knyttet til århundrene 18-, 19-, og 20- og er i den forstand *brukt opp*. Koding av århundre gjennom valg av eksplisitte verdier for andre I_j er på samme måte ikke mulig å gjennomføre. Men det finnes én måte å omgå dette på:
- For alle gamle/eksisterende fødselsnumre er kontrollsummen S_1 assosiert med kontrollsymbolet K_1 pr. definisjon lik 0 modulo 11.
- Det vil si at for hver av de nye identifikatorverdiene som blir tilgjengelig, er S_1 *ulik* 0 modulo 11, det vil si S_1 kan anta verdiene 1, 2, ..., 10 modulo 11.
- Hver serie med fast verdi av S_1 er en mengde av identifikatorverdier som har samme egenskaper som det opprinnelige fødselsnummersystemet. Det vil si at for hver dato på formen DDMMÅÅ er det tilgjengelig minst 826 løpenumrerverdier. Feildeteksjonsegenskapene *innenfor* hver slik klasse er også identisk med egenskapene i det opprinnelige systemet. Deteksjon av feil på tvers av klasser vil være avhengig først og fremst av kontrollsymbolet K_2 , men dersom vi velger koding for århundre som beskrevet ovenfor, vil også symbolet i ”posisjon K_1 ” fremdeles bidra til forbedret feildeteksjon.
- Dersom dette alternativet blir valgt, vil vi derfor anbefale å starte tildeling av numre fra serien med $S_1 = 1$.
 - **Kapasitet:** Under forutsetning av at behovet for individnummer pr. dag holder seg under 826 pr. dag (kfr. Figur A5.2), vil dette systemet holde et helt århundre for $S_1 = 1$, og så lenge disse forutsetningene er oppfylt videre vil systemet kunne vare i 1000 år. Dersom behovet overstiger denne kapasiteten kan man bruke to verdier av S_1 pr. århundre.
 - **Koding for århundre** vil dermed være gitt direkte gjennom verdien av S_1 .
 - **Feilkontroll** er bestemt først og fremst av K_2 , men med et viktig ekstra bidrag fra K_1 fordi på et gitt tidspunkt i århundre x vil de fleste levende personer ha verdier av S_1 som er begrenset til x eller $x+1$. Å beregne den eksakte effekten av dette i forhold til robusthet mot feil krever en mer detaljert analyse. Man kan med en viss rett si at vi beholder ”1.7 kontrollsiffer” istedenfor de opprinnelige 2. (Dette gjelder bare dersom århundre kodes som beskrevet her.)
 - **Kjønn** kan kodes som før.

Denne løsningen kan kombineres med alternativ 0X eller alternativ 6 hvis det er nødvendig for å utvide kapasiteten ytterligere.

Sannsynlighet for uoppdaget feil, P_{ud} , er vist i Figur 1 som en funksjon av enkeltsymbolfeil-sannsynlighet p (blå prikket kurve.) Denne er tegnet under forutsetning av at opprinnelige identifikatorverdier ($S_1 = 0$) og nye verdier med $S_1 = 1$ eksisterer samtidig. Som vist blir påliteligheten svekket med en faktor mellom to og tre i forhold til eksisterende fødselsnummer. Dersom bare $S_1 = 1$ eksisterer blir feilsannsynligheten som for opprinnelig fødselsnummer. Dersom bare to typer identifikatorverdier som svarer til henholdsvis $S_1 = i$ og $S_1 = i + 1$ eksisterer, får man igjen en feilsannsynlighet tilsvarende den blå prikkete kurven.

Figur 1 viser også sannsynligheten P_{AFB} , sannsynligheten for at en Aksepterer Feile identifikatorer som er Brukt. Som en tilnærming antar vi at for de fleste dager (DDMMÅÅ) i dagens fødselsnummerordning er mindre enn halvparten av de gyldige identifikatorverdiene faktisk tildelt. Det gir derfor mening å starte tildeling av identifikatorverdier i serien med $S_1 = 1$ fra «andre enden». Som en ytterligere tilnærming *definerer vi* P_{AFB} til å være sannsynligheten for feil ved Alternativ 1 under forutsetning av at løpenummer for seriene for henholdsvis $S_1 = 0$ og $S_1 =$

1 er ulike. Som Figur 1 viser vil dette forbedre påliteligheten i Alternativ 1, som da vil ligge nærmere dagens fødselsnummer.

Alt. 5. Skala for måned utvides

Tallserie 21-32 for måned tas i bruk fra f.eks. 2040. Senere kan seriene 41-52, 61-72, 81-92 benyttes. Kan gjøre skifte ved hvert århundre. (Passe på at det ikke kan forveksles med andre nøkler).

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.5 Skala måned DDMMÅÅIIKK	Tja	0	+		des		+	+	-	

- Skala for måned utvides (med 21-32, 41-52 etc)
 - *Nominelt en femdobling av kapasitet*
- Mer effektivt: Legge et visst antall *12 til faktisk måned: 13-24, 25-36, osv.
 - *Nominelt en åttedobling av kapasitet*
- Uansett mindre effektivt enn alternativ 1
- Dette betyr at gamle numre ikke trenger å endres.
- Feilkontroll beholdes som idag.
- Kapasiteten blir max 826 pr. dag.
 - Men avhengig av praktisk implementasjon
 - Skifte ved århundre endrer ikke dette
- Tilstrekkelig for tallene i befolkningstilvekst frem til 2100 der maksimale pr. dag. er 738 (men liten margin)
- Fare for konflikt med andre nummerserier.
- ”Informasjon” i dato blir besudlet.

Alt. 6. 14 sifre

Århundre får 2 sifre (4 sifre for år) og 4 sifre for individ. Alternativt kan 13 posisjoner benyttes hvis ett kontrollsiffer er tilfredsstillende.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.6 14 sifre DDMMÅÅÅÅIIKK	Ja	++	0	--	des			+	-	

14 sifre (eventuelt 13 sifre med et kontrollsiffer)

Feilkontroll modulo 17

Med 14 sifre

- Kapasitet 3460 (pr. dag)
- Oppdager alle enkle feil og alle doble og mange flere.
- Rikelig kapasitet.

Med 13 sifre

- Kapasiteten blir 5882 pr. dag.
- Rikelig kapasitet men langt!

Informasjonsløs/begrenset informasjon samt numeriske tegn:

Alt. 7. Dag inngår i individnummer, svensk løsning, 11 sifre og 2 kontrollsifre.

Dette alternativ ligner på den nye svenske løsningen, men tilpasset formatet på dagens norske fødselsnummer. Får noe redusert informasjon da dag (DD) går over til å være individnummer (II). Må her unngå å bruke eksisterende fødselsnumre eller andre viktige nummerserier.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.7 Svensk løsning IIMMÅÅIIKK	Tja	+	+		des	0	+	+	-	

- *Dag inngår i individnummer, svensk løsning, 11 sifre, derav 2 kontrollsifre*
- Ca tredobling av dagens kapasitet
 - 82645 pr MMÅÅ (gjennomsnitt pr dag 2660) dersom man ikke korregerer for eksisterende serier
 - Følsom for *måned*svariasjon og ikke for *dag*svariasjon (dette er bedre)
 - Kapasitet antagelig robust, men ikke så robust som Alt. 1 eller alt. 2
- Konflikt med eksisterende fødselsnummer og med D-nummer
 - Mister ca. 62 % av kapasitet når vi tar bort eksisterende nummerserier
 - Restkapasitet ca. 31400 pr MMÅÅ, eller gjennomsnitt 1013 pr. dag
 - Kan holde til år 2150, men marginalt
- Samme feilkontroll som i dagens fødselsnummer
- **Semantikk forsvinner**; Vanskeligere å huske?
- Programvareproblem (test på dato)

- Koding for århundre kan fungere gjennom å ta i bruk ”de nye” datoene (32-40, 72-99) i dagsfelt)
- Variasjon: DDIIÅÅIIKK
 - gir kapasitet 82645 pr “månedsdag” eller 6887 pr DD..ÅÅ
- Variasjon: IIIIÅÅIIKK
 - gir kapasitet 8264463 pr “år” eller 22580 pr DDMMÅÅ

Alt. 3. 11 sifre, 2 kontrollsifre

Kan karakteriseres som et slags ”kundennummer” uten informasjon. Starter med lave numre med 0’er foran.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.3 Kundennummer IIIIIIIIKK	Nei	++	+	-	des	-	-	+	-	

- Kundennummer 11 sifre, 2 kontrollsifre
- Kapasitet “22580 per dag over hundre år” – men bare en tredjedel av f. eks Alt. 4 med løpenummerlengde 7 og $q=23$, og med betydelig verre feildeteksjonsegenskaper enn Alt. 4
- Samme feilkontroll som i dagens fødselsnummer
- Mulig konflikt med D-nummer
- Semantikk forsvinner;
- Programvareproblem (test på dato)

Sannsynlighet for uoppdaget feil er vist i Figur 1 (heltrukken svart kurve.) Denne sannsynligheten gjelder selv om det dagens fødselsnummer og D-serien er i bruk parallelt.

Alfanumerisk og informasjonsbærende:

Alt. 0X. 11 posisjoner, 2 kontrollsifre. Tillater ”siffer X” = 10

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt. 0X	Tja	-	+		11	+	-	+	0	

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
DDMMÅÅXXXXKK										

- Øker alfabetet fra $A_{10} = \{0,1,2,3,4,5,6,7,8,9\}$ til $A_{11} = \{0,1,2,3,4,5,6,7,8,9,X\}$.
- Kan bruke samme moduluskontroll som før, under regelen $X=10$
- Kapasitet 1331 pr dag minus det som er brukt fra før – det vil si at kapasiteten vil
 - øke med 505 pr dag i perioden frem til 2054
 - være ca. 505 +165 (se Tabell 1 og 2) i perioden 2054-2099
- **Koding for århundre er problematisk ; ingen opplagte løsninger på dette**
- Kan fjerne et kontrollsymbol og øke kapasitet til 14641 per dag, minus det som er brukt (tilsv. Alternativ 1)
- Enkel tilpasning til dagens identifikator, kan være midlertidig kriseløsning

Alt. 2. 11 posisjoner: Bruke alfabetutvidelse

Må også utvide moduluskontroll.

Variant 1: DDMMÅÅ IIIBK

11 posisjoner, 1 kontrollsifre. Frigjort K bokstav som angir århundre. Bokstav i individdelen.

- F.eks A for 2000, B for 2100, osv.
- Kapasitet 909 pr, dag pr. bokstav
- Kapasitet rikelig, men lite effektiv utvidelse (sammenlignet med versjon 2 nedenfor)
- **Må skifte moduluskontroll**
- **Kompatibilitet og sameksistens med gammelt fødselsnummer: Lite robust mot feil som forandrer gammel fødselsnummer til ny identifikator og vice versa**

Variant 2: DDMMÅÅAAAKK, der de siste fem symbolene er fra et alfabet med q symboler. **Minst tre av disse må være bokstaver/ikkedesimale**, for feildeteksjon i sameksistens med opprinnelig identifikator. Når vi diskuterer Alternativ 2 i oppsummering, er det denne variant 2 vi refererer til.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.2 Variant 2 DDMMÅÅAAKK	Ja	++	++		<i>q</i>	++	+	+	-	+

Koding av århundre og kjønn: Kan gjøres på flere måter, for eksempel gjennom bruk av første bokstav.

For hvert eksempel nedenfor er det gitt to lister av vektorer for to kontrollsymboler, henholdsvis K_1 og K_2 . Dersom vi bare skal bruke ett kontrollsymbol, det vil si Alternativ 2 i kombinasjon med alternativ 1, brukes bare K_1 .

Det finnes mange andre kodekonstruksjoner og måter forkorte på; så mange at vi ikke har gjort noe fullstendig søk eller noe grundig forsøk på optimalisering.

«Kapasitet» svarer til minimumverdi (minimert over alle datoer, det vil si minimert over alle partielle kontrollsummer. Kritisk paritetsverdi er en teknisk parameter som identifiserer den partielle kontrollsummen) av *antall kodeord pr.dag*, under betingelsen at de fem siste symbolene inneholder høyst et gitt max antall desimale symbol.

- Dersom koden skal eksistere sammen med den gamle Selmerkoden, men med forskjellig feilkontrollmekanisme, kan vi ikke tillate bare desimale symboler i et kodeord. Rekkene i tabellen som tillater fem desimale symboler gir altså de største kodene, men tillater ikke sameksistens med Selmerkoden.
- Dersom vi tillater ned til én bokstav (max fire desimale symbol) kan vi skille mellom gammel og ny kode på en støyfri kanal (men på en støyfri kanal trenger vi ikke kontrollsymbol,) men vi kan ikke detektere enkle feil som skifter mellom fire og fem desimale symbol. Derfor er rekkene med fire og fem desimale symbol markert med grå bakgrunn.
- Dersom vi har to kontrollsymbol og vi tillater max to desimale symbol i de siste fem symbolene (dvs minst tre bokstaver), kan den nye koden detektere to feil (og en del andre sannsynlige feil) i sameksistens med den gamle Selmerkoden. Disse verdiene er derfor markert med **blått**. Dersom vi bare har ett kontrollsymbol (Alt. 2 sammen med alt. 1) kan vi tolerere opptil tre desimale symbol (minst to bokstaver) og likevel detektere alle enkle feil (og alle enkle transposisjoner), og disse verdiene er også markert med **blått**.

For eksempel: For $q=23$ med to kontrollsymbol, ønsker vi å se på løpenumre som oppfyller betingelsen at de fem siste symbolene har max to desimale siffer, det vil si minst tre bokstaver. Denne "koden" vil detektere alle kombinasjoner av feil i to posisjoner. For hver dag er det tilgjengelig minst **7535** slike løpenumre, noen dager har litt mer. Dersom vi bare ønsker ett kontrollsymbol velger vi bare vektene som svarer til K_1 og maksimum tre desimale tall (eller ekvivalent, minst to bokstaver) av de siste fem, det finnes minst **247229** slike kodeord for hver dag. Denne "koden" vil detektere alle enkle feil, og en del andre feil.

Dette er altså opptellinger av kodeord, og ikke en probabilistisk analyse. Det er også litt variasjon over de forskjellige datoene, det som er gitt i tabellene er minimumverdier, men de ligger svært nær gjennomsnittsverdiene (som vil være det optimale.)

Koding av århundre kan gjøres på flere måter, for eksempel gjennom valg av første bokstav. Hvert valg av bokstav gir ca 580 identifikatorer pr. DDMMÅÅ, og ved å bruke to bokstaver pr. århundre kan man da kode for både kjønn og århundre. Dersom behovet pr kjønn og dag holder seg under 580, vil dette systemet vare i seks hundre år.

$q = 17$. K_1 , vekter: 3 9 10 13 5 15 11 16 14 6 1

$q = 17$. K_2 , vekter: 7 16 3 12 12 7 15 8 14 1 0

	To kontrollsymbol	Ett kontrollsymbol
Max antall desimale symbol	Kapasitet	Kapasitet
0	55	981
1	464	8013
2	1645	28181
3	3342	57013
4	4561	77631
5	4913	83521

$q = 19$. K_1 , vekter: 2 4 8 16 13 7 14 9 18 10 1

$q = 19$. K_2 , vekter: 11 9 4 9 3 3 13 16 11 1 0

	To kontrollsymbol	Ett kontrollsymbol
Max antall desimale symbol	Kapasitet	Kapasitet
0	158	3106
1	1059	20361
2	3087	58731
3	5322	101361
4	6575	125056
5	6859	130321

$q = 23$. K_1 , vekter: 5 2 10 4 20 8 17 16 11 14 1

$q = 23$. K_2 , vekter: 22 16 7 4 4 11 14 11 6 1 0

Ett kontrollsymbol

To kontrollsymbol

Max antall desimale symbol	Kapasitet	Kapasitet
0	693	16142
1	3387	78229
2	7535	173747
3	10739	247229
4	11973	275492
5	12167	279841

Sannsynlighet for uoppdaget feil er vist i Figur 1 (heltrukken grønn kurve.) Denne forutsetter vektor henholdsvis 9, 22, 13, 5, 21, 15, 10, 6, 3, 1, 0 og 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, men mange andre løsninger eksisterer – vi har ikke gjort forsøk på å optimalisere. Figur 1 viser også sannsynlighet for at identifikatorverdier av type Alternativ 2 blir mistolket som et gyldig eksisterende fødselsnummer (sort prikket linje) eller omvendt (turkis prikket linje.) Vi ser at disse sannsynlighetene ikke er signifikante i forhold til total feilsannsynlighet.

Alt. 8. 11 posisjoner, 2 kontrollsifre.

Angir århundre i individdelen. Første individposisjon bokstav for århundre 2000, andre individposisjon for århundre 2100, tredje 2200, første og andre 2300, første og tredje 2400, andre og tredje 2500, tre bokstaver 2600.

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt. 8 DDMMÅÅAIIKK	Ja	--	-		q	-	-	+	-	-

- **Format:** DDDMMM AII KK
- **Kapasitet:** $(q-10) \cdot 100 \cdot (10/q)^2$ pr dag (første to århundre...), der q er alfabetstørrelsen.
- For liten kapasitet:

q	11	13	17	19	23	29	31	37
Kapasitet	83	178	242	249	246	226	219	197

Alfanumerisk og informasjonsløst/begrenset informasjon

Alt. 4. Alfanumerisk, m informasjonsposisjoner

	1	2	3	4	5	6	7	8	9	10
	Info	Kap	Feil	Len	Alf	Ko	Utv	En	Pr	Tek
Alt.4 AAAAAAK(K)	Nei	++	++	+	q	0	++	+	-	++

Vi antar m løpenummerposisjoner med k kontrollsifre ($k = 1$ eller 2). Alle symboler er hentet fra et alfabet $\mathcal{A}$ med $s(\mathcal{A})$ symboler, der $s(\mathcal{A})$ er lik et primtall q . Kapasiteten, som kan jevnføres med behovet (frem til år 2100) i Figur A5.1 i Appendiks 5, er gitt i Tabell 6:

q	5 infopos.	Varer	6 infopos.	Varer	7 infopos.	Varer	8 infopos.	Varer
11					19487171	Til 2086	214358881	1000år???
13			4826809	For lite	62748517	250 år?	815730721	
17			24137569	Til 2100	410338673		6975757441	Eksport til Kina
19			47045881	200 år?	893871739		1,6984E+10	Global eksport
23	6436343	For lite	148035889		3404825447	Eksport til Kina		
29	20511149	Til 2088	594823321		1,725E+10	Global eksport		
31	28629151		887503681					
37	69343957		2565726409					

Tabell 6

Varighet i tabellen er beregnet ut fra høy nasjonal vekst (Se Figur A5.1), og umiddelbar tildeling av 5 millioner identifikatorer (pr. 19.03.2012!). Etter 2100 har vi ekstrapolert resultatene fra Figur A5.1, og disse ekstrapolerte prognosene er selvsagt svært usikre. Vi forutsetter videre at ikke andre grupper blir tildelt identifikatorer. Hvis andre grupper blir tildelt identifikatorer, blir varigheten kortere.

Vår anbefaling, 6 informasjonsposisjoner (løpenummersymboler) og $q=23$ er markert med gult i tabellen. Dette gir ca. 148 millioner tilgjengelige identifikatorer.

Utvidbarhet: Ved å velge en fast og stor nok q , kan eventuelt identifikatoren velges med en liten m i begynnelsen. Ettersom behovet vokser, kan m økes for å utvide kapasiteten (og de ekstra symbolene antas å være 0 for identifikatorer som allerede er tildelt.)

Antall kontrollsymboler k velges etter behov for identifikatoren. Ved spesielle behov kan man bruke ekstra kontrollsymboler (det vil si bruke en større verdi for k .) Dette går bra så lenge $m+k \leq q$. En alternativ metode for å legge til rette for fremtidig utvidbarhet er å starte med et mindre alfabet, og utvide alfabetstørrelsen ved behov (inntil $s(\mathcal{A}) = q$.) Vi antar at ved normal bruk er det nok å bruke $k = 1$ eller $k = 2$. Fordeler/ulempene ved disse to alternativene er diskutert mer generelt andre steder i dette dokumentet, og valg av antall kontrollsymboler må foretas på strategisk grunnlag: $k = 1$ er mye enklere, $k = 2$ gir mye bedre pålitelighet.

Ny teknologi: Dersom nye teknologi stiller krav til bruk av en mye større mengde av mulige identifikatorverdier, for eksempel for å assosiere biometriske verdier til identifikatoren eller for å lage mer "luft" mellom mulige identifikatorverdier, er dette lett å oppnå med dette alternativet fordi utvidbarhet er lett å oppnå.

Sannsynlighet for uoppdaget feil er vist i Figur 1 (heltrukken rød kurve for to kontrollsymboler, rød prikket linje for ett kontrollsymbol.) Denne forutsetter vektor henholdsvis

9, 22, 13, 5, 21, 15, 10, 6 og 11, 10, 9, 8, 7, 6, 5, 4, men mange andre løsninger eksisterer – vi har ikke gjort forsøk på å optimalisere.

Vi kan konstatere at bruk av ett kontrollsymbol vil gi en betydelig redusert pålitelighet i forhold til andre alternativ, og vi **anbefaler derfor to kontrollsymbol og total lengde 8**. Lengden kan om ønskelig trivielt økes til 11 ved å legge til tre nullere i fast valgte posisjoner.

Appendiks 4: Noen løsninger i andre land

Vi har studert hvor mange kontrollsifre som benyttes i lignende systemer varierer fra land til land. I de fleste tilfeller brukes ett kontrollsiffer og i de øvrige tilfeller benyttes enten 2 eller 0 kontrollsifre. Ikke alle land har et system som er landsdekkende eller obligatorisk for alle borgere så sammenligningen er ikke alltid helt relevant men gir en god indikasjon. Det er litt overraskende at mindre land som Norge og Albania bruker to sifre mens for eksempel USA har ingen kontrollsifre i sitt Social Security Number SSN.

Det er kanskje verdt å merke seg at Danmark som tildelte et 10-sifret personnummer med (3-sifret individnummer) og opprinnelig 1 kontrollsiffer har besluttet å tildele numre uten kontrollsiffer når det ikke er tilgjengelig numre med kontrollsiffer. Dette betyr at feilkontrollen mellom numrene er sterkt redusert.

I noen land økes kapasiteten i personnumre ved å modifisere DDMM i fødselsinformasjon for eksempel ved å la DDMM svare til ikke gyldig datoer eller måneder.

Appendiks 5: Krav til kapasitet

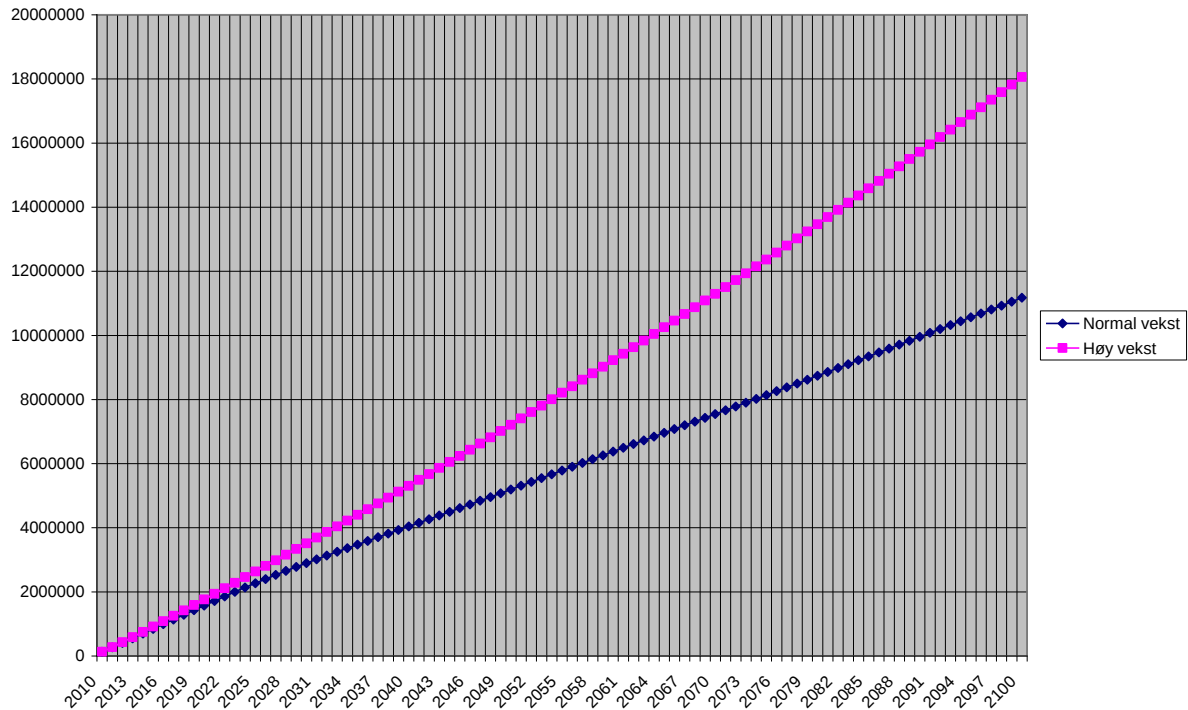
Det er vanskelig å lage pålitelige prognoser for hvor mange identifikatorer som er nødvendig for en gitt dag eller periode: dette tallet avhenger av (i) biologiske fødselsprosesser, som avhenger av blant annet befolkningsstørrelse og –alderssammensetning, og er vanskelig å forutse, (ii) innvandring, som er avhengig av globale økonomiske og politiske forhold, og er vanskelig å forutse, (iii) politiske beslutninger i forhold til hvilke grupper av individer som skal tildeles identifikatorer (bare statsborgere? Bosatte? Asylsøkere? Skattebetalere? Korttidsbesøkende?), som er umulig å forutse.

Statistisk Sentralbyrå har utarbeidet prognoser for behov for nye identifikatorverdier, basert på forskjellige forutsetninger (lav; normal; høy vekst med hensyn på fødsler og immigrasjon).

Informasjonsløse identifikatorer: Det kumulative behovet, som er relevant for en informasjonsløs identifikator er oppsummert i **Figur A5.1**.

- Det viser at behovet for **nye** identifikatorverdier ved høy vekst er ventet å passere 4 millioner i 2033, 8 millioner i 2054, 12 millioner i 2074, og 18 millioner i 2100.
- Dersom eksisterende befolkning også skal tildeles en slik informasjonsløs identifikatorverdi, øker behovet med ca 5 millioner.
- Dersom regelendringer medfører at andre grupper (som ikke i dag får tildelt fødselsnummer) får tildelt en identifikator, kan behovet øke betydelig. Dette gjelder personer som søker arbeidstillatelse, personer som omfattes av EØS-registrering, asylsøkere, og lignende. Noen av disse får i dag tildelt identifikatorer, men ikke alle. For de neste årene (2012-2014) er antallet i disse gruppene årlig i størrelsesorden 75.000-90.000 (se [Olsen]), men det er betydelig usikkerhet knyttet til disse tallene. Det er også

grunn til å tro at det forekommer dobbeltelling her. Uansett er det vanskelig å forutse betydningen av denne gruppen i forhold til kapasitetsbehov for personidentifikatoren, men i et *hundreårsperspektiv* er det ikke urimelig å anslå at denne gruppen vil øke kapasitetsbehovet med ti millioner eller mer.



Figur A5.1: Kumulativt behov for **nye** identifikatorer, inkludert fødsler og immigrasjon

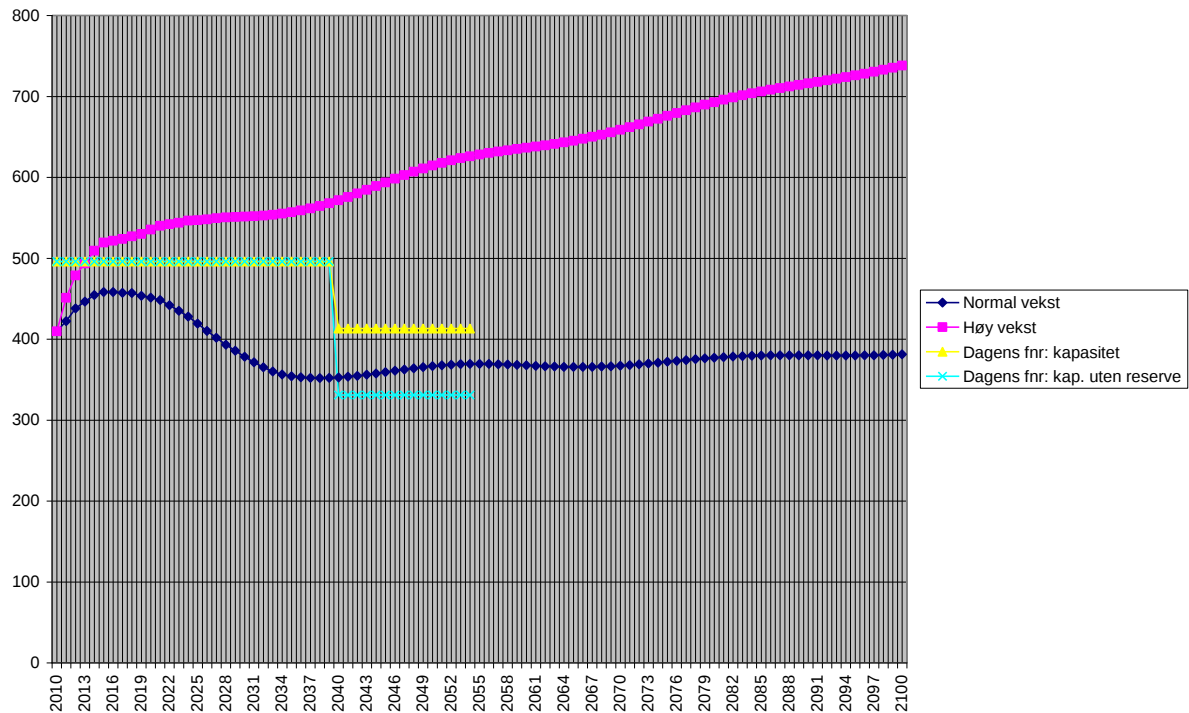
Identifikatorer med innbakt dato og annen informasjon: For en identifikator som inkluderer informasjon om fødselsdag, er det kritiske kapasitetsbehovet bestemt ved hvor mange som blir født hver enkelt dag. Prognosene baserer seg på generelle forventninger om befolkningsvekst.

I tillegg er antall fødsler normalt ikke jevnt fordelt over årets dager. Fødselstall er til en viss grad sesong betont. Dessuten er det ikke vanlig å sette i gang fødsler eller å foreta planlagte keisersnitt i helger og høytider. Analysene fra statistisk sentralbyrå tar utgangspunkt i at på årets ”travleste” blir det født et antall individ som overstiger årsgjennomsnittet med 14%. Anslag for antall individer født på årets travleste dag er vist **Figur A5.2**.

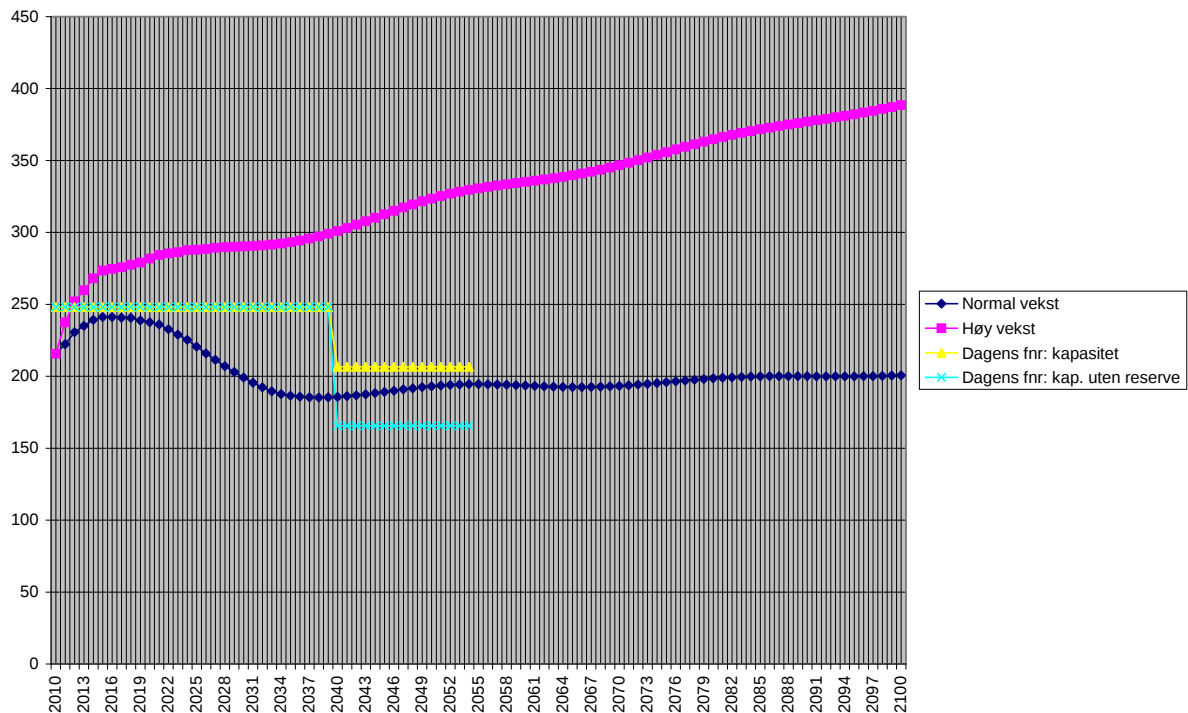
- Det viser at kapasitetsbehovet pr. dag ved høy vekst er ventet å passere 500 i 2014, 600 i 2047, og 700 i 2083.
- Ved normal vekst er det ventet at kapasitetsbehovet pr. dag når en topp på i underkant av 460 rundt år 2015, og at det fra år 2028 vil stabilisere seg på verdier under 400 pr. dag.
- Dersom vi deler populasjonen inn i kjønn, vil den relative variansen med hensyn til antall individer født pr. dag innenfor hvert kjønn være større enn for populasjonen som helhet. I **Figur A5.3** viser vi hvordan dette slår ut når kjønn er den del av informasjonen som kodes inn i identifikatoren. Vi antar at halveringen av gruppen medfører en økning av variansen med en faktor på 1.4 (kvadratroten av 2), slik at på årets travleste dag kan det bli født 20% flere enn gjennomsnittet pr. kjønn.⁵

⁵ På den annen side er dette kanskje litt for konservativt, ettersom variansen i fødselstall over året ikke bare skyldes tilfeldige prosesser, men også systematiske forskjeller mellom ukedagene.

- Dersom regelendringer medfører at andre grupper som i dag ikke får tildelt fødselsnummer vil få det i fremtiden, vil dette som nevnt ovenfor øke kapasitetsbehovet betydelig. Tallene fra Utlendingsdirektoratet (se [Olsen indikerer at i *verste fall* kan vi snakke om dagsbehov som øker med 200-250, bare for de første årene (2012-2013). Disse tallene inkluderer imidlertid som sagt mange dobbelttellinger. I praksis løses disse sakene i dag av og til ved at innvandrere blir tildelt en vilkårlig «vedtaksfødselsdag» for å fordele bruk av fødselsnumre.



Figur A5.2: Anslag for antall individer født på årets tetteste befolkete fødselsdag, under forutsetninger om henholdsvis normal og høy nasjonal vekst. "Andre grupper" som i dag ikke får tildelt identifikator men som i fremtiden kan få det, er ikke tegnet inn. Også kapasiteten i dagens fødselsnummersystem er tegnet inn, med og uten reserveserien (9-serien).



Figur A5.3: Anslag for antall individer pr. kjønn født på årets tetteste befolkete fødselsdag, under forutsetninger om henholdsvis normal og høy nasjonal vekst. "Andre grupper" som i dag ikke får tildelt identifikator men som i fremtiden kan få det, er ikke tegnet inn. Også kapasiteten i dagens fødselsnummersystem er tegnet inn, med og uten reserveserien (9-serien).

Usikkerhet 1. Figurene ovenfor er beregnet på grunnlag av regnearket «Fremskrevet behov for nye fødselsnummer per år og dag», som er gjort tilgjengelig for oss fra Statistisk Sentralbyrå. Nyere tall fra Statistisk Sentralbyrå <http://www.ssb.no/emner/02/02/folkendrkv/> viser at de faktiske fødselstallene for 2011 var 60.200, eller 2200 *lavere* enn prognosen for normal vekst. Tallene for innvandring ser ut til å være litt *høyere* enn normal vekst skulle tilsi, slik at det faktiske samlede behovet for nye identifikatorer for 2011 ligger *midt mellom normalvekst-prognosene og høy vekst-prognosene*.

Usikkerhet 2 - og svarte svaner. Vi vil videre kommentere at prognosene er følsomme for svingninger i svært komplekse samfunns- og økonomisvingninger. Nassim Taleb (Taleb, 2007) har innført begrepet "svarte svaner" om hendelser som blir ansett for å være svært usannsynlige fordi de inntreffer som følge av forhold som ikke er godt bygget inn i de teoretiske modellene vi til enhver tid bruker for å beskrive den verden vi lever i. *Slike hendelser blir derfor feilaktig ansett for å være usannsynlige* før de faktisk inntreffer, men kan ha betydelig innvirkning på samfunnsutviklingen.

Ettersom vi kan være rimelig sikker på at svarte svaner eksisterer, og på at uforutsette hendelser *vil* inntreffe, er vi ukomfortable med å anbefale personidentifikatorsystemer som har en kapasitet som ligger marginalt i overkant av anbefalt behov. Vi foreslår istedenfor en konservativ holdning som innebærer å bruke identifikatorer som (1) har en romslig kapasitet i forhold til anslåtte behov, og som (2) ved enkle grep er utvidbare.

Referanser

Selmer, E. S. (1964). Personnummerering i Norge: Litt anvendt tallteori og psykologi. *Nordisk Matematisk Tidskrift*, 36-44.

Skattedirektoratet. (2012). Forespørsel: Ny personidentifikator i Folkeregisteret, Analyse av kapasitet og egenskaper.

J. Verhoeff, "Error Detecting Decimal Codes", Mathematical Centre Tract 29, The Mathematical Centre, Amsterdam, 1969.

Torleiv Kløve, "Codes for Error Detection", World Scientific, 2007.

Grunndatarapporten,

http://www.regjeringen.no/nb/dep/fin/dok/rapporter_planer/rapporter/2007/utveksling-av-grunndata-pa-personinforma.html?id=491849.

Statistisk Sentralbyrå, "Fremskrevet behov for nye fødselsnummer per år og dag" (xls-fil, vedlegg til oppdrag)

Nassim Nicholas Taleb, 'The Black Swan: The Impact of the Highly Improbable,' 2007

Vigdis M. Olsen...-Prognoser for personer som registreres i UDB for første gang...

Rapport om egenskaper ved ny personidentifikator: Del 2

Tor Helleseeth og Øyvind Ytrehus
Universitetet i Bergen/Selmersenteret
Revidert 04.04.2013

Innhold

Innledning.....	1
1) Alternativ 1 og Grunndatarapportens anbefaling	1
2) Alfabet og alfabetstørrelse.....	2
3) Kompatibilitet og sameksistens med dagens identifikator	10
4) Kommentar vedrørende Alternativ 7.....	12
Referanser	13

Innledning

Denne rapporten er en fortsettelse av ” **Rapport om egenskaper ved ny personidentifikator**”, datert 23.04.2012 [1]. Notasjoner og begreper er innført og definert i det dokumentet, som vi forutsetter lest.

I *denne* rapporten tar vi opp:

- 1) En avklaring av vårt anbefalte Alternativ 1 kontra Grunndatarapportens [2] anbefaling.
- 2) Tegnsystem.
- 3) Kompatibilitet og sameksistens med nåværende fødselsnummersystem, spesielt for anbefalte alternativ
- 4) En kommentar til en kommentar om Alternativ 7.

1) Alternativ 1 og Grunndatarapportens anbefaling

Grunndatarapportens [2] subsidiære forslag er på formen $DDMM\hat{A}I_1I_2I_3I_4K_2$, mens de gamle/eksisterende identifikatorene er på formen $DDMM\hat{A}I_1I_2I_3K_1K_2$. Vi forutsetter at Grunndatarapporten la til grunn samme feilkontroll på det gjenværende kontrollsymbolet (K_2) som i dag. Kapasiteten for hver $DDMM\hat{A}$ i en identifikator på formen $DDMM\hat{A}I_1I_2I_3I_4K_2$ vil være ca 9090, men av disse er en nummerserie på ca. 826 brukt opp (for praktiske formål) i dagens fødselsnummer, så gjenværende kapasitet er ca. 8264.

- Utfordringen med grunndatarapportens forslag er (a) å velge $I_1I_2I_3I_4$ på en slik måte at man ikke ved et uhell genererer en identifikator som er brukt før, og (b) å kode for århundre. (Koding av århundre skjer i dagens fødselsnummer gjennom bruk av første siffer I_1 , som er knyttet til århundrene 18-, 19-, og 20- og som i den forstand er *brukt opp*. Koding av århundre gjennom valg av eksplisitte verdier for andre I_j er på samme måte ikke mulig å gjennomføre.)

Vi kan ikke se hvordan Grunndatarapportens subsidiære forslag slik det er definert og uten presiseringer, skal kunne benyttes som personidentifikator med krav om blant annet koding av århundre, og anbefales derfor ikke. Derimot er det mulig å beskrive en mer presis struktur i bruk av kontrollsymbolene, som gjør det mulig å benytte identifikatoren.

Det som presist kjennetegner identifikatorverdier i *eksisterende* fødselsnummerserie, i forhold til Alternativ 1, er at kontrollsymbolet K_1 er valgt slik at den tilhørende kontrollsummen S_1 *er lik* 0 modulo 11. En systematisk måte å unngå ”gamle” verdier er derfor i fremtiden å velge I_4 ($=K_1$) slik at den tilhørende kontrollsummen S_1 *er ulik* 0 modulo 11.

For i tillegg å kunne kode for århundre på en systematisk måte, samt å beholde en viss feildeteksjonsevne, foreslår vi altså i første omgang å tildele verdier i en serie definert ved at I_4 ($=K_1$) settes slik at den tilhørende kontrollsummen S_1 *er lik* 1 modulo 11. Når denne serien er brukt opp kan en fortsette med å velge I_4 ($=K_1$) slik at S_1 *er lik* 2 modulo 11, og så videre. Hver av disse seriene har 826 identifikatorer pr. DDMMÅÅ; til sammen har vi derfor 8260 ubrukte identifikatorer for hver DDMMÅÅ. Dersom dagens langtidsprognoser for årlig befolkningsvekst er gyldig «for all fremtid» (noe som høres urimelig ut!), vil Alternativ 1 derfor kunne brukes i tusen år, og med implisitt koding for århundre.

Bruk av identifikatoren

Personer som har fått tildelt en identifikator etter dagens fødselsnummerordning, beholder denne. På et eller annet tidspunkt vil det ikke være flere tilgjengelige identifikatorer etter gammel ordning, og nye personer vil få tildelt et nummer fra serien med $S_1 = 1$.

Identifikatoren er, som i dagens fødselsnummer, unik, og vil primært brukes som en indeks for tilgang til persondata fra en database. Dersom man er i en «offline»-situasjon, men har tilgang til elektronisk registreringsutstyr, og ønsker å hente ut de opplysningene som, utover fødselsdato, ligger implisitt i identifikatoren (det vil si kjønn og århundre), er det en enkel sak å modifisere programvaren som sjekker kontrollsymbolene slik den kan trekke ut kjønn og århundre også. Kjønn angis som i dag med like tall for kvinner og ulike tall for menn og kan leses direkte ut for de som kjenner denne definisjonen.

Langsiktig og fremtidig opsjon

Hvis det på et fremtidig tidspunkt skulle være ønskelig å utvide kapasiteten ytterligere, er det mulig hvis det aksepteres en noe svakere feilkontroll. I_4 inngår i individnummeret og benyttes ikke som kontrollsiffer (DDMMÅÅ $I_1I_2I_3I_4K_2$). K_2 defineres slik at den tilhørende kontrollsummen S_2 *er lik* 1 modulo 11. Når denne serien er brukt opp kan en fortsette med å velge K_2 slik at S_2 *er lik* 2 modulo 11, og så videre. Hver verdi av S_2 *ulik* 0 gir en kapasitet på 826 9090 identifikatorverdier per dag. En slik strategi vil svekke feilkontrollen betydelig.

2) Alfabet og alfabetstørrelse

To grunnleggende valg i design av et identifikatorsystem er (a) valg av alfabetstørrelse og (b) valg av alfabet. Disse valgene henholdsvis (a) bestemmer kapasitet med hensyn på antall tilgjengelige numre for tildeling (i stor grad), og (b) påvirker systemets brukervennlighet og robusthet. Siden det valgte alfabet bør ha riktig størrelse, bør valg (a) fortrinnsvis gjøres før valg (b), men noen egenskaper ved for eksempel utvalget av tilgjengelige symboler i et alfabet tilsier at valget (a) ikke er helt uavhengig av valget (b).

Valg av alfabetstørrelse

Som beskrevet i vår tidligere rapport [1] bør alfabetstørrelsen være en primtallspotens, og helst et primtall, med tanke på forenklet feilkontroll. Med en gang alfabetstørrelse og identifikatorsystem er bestemt, er også kapasiteten bestemt – det vil si at valg av selve alfabetet ikke påvirker kapasiteten, slik vi tidligere har definert dette begrepet.

Valg av alfabet

Med begrepet *alfabet* mener vi den konkrete mengden av symboler som er tillatt i identifikatoren. I dagens fødselsnummer består for eksempel alfabetet av all desimalssiffer, det vil si at alfabetet er $\{0,1,2,3,4,5,6,7,8,9\}$.

Vi vil se på alfabeter som består av desimalsifre i tillegg til en mengde av utvalgte bokstaver. Med «bokstav» mener vi (store) bokstaver fra «A» til og med «Z». Vi vil ikke ta med «E», «Ø» og «Å», fordi disse ikke er tilgjengelig på alle tastatur som brukes i verden, og disse bokstavene er også grafisk like andre bokstaver.

Man kan tenke seg identifikatorer som har forskjellige alfabetet for de forskjellige posisjoner (for eksempel «desimaltall i de første ti posisjoner etterfulgt av én bokstav i posisjon 11»). Slike systemer brukes i noen land, og kan motiveres ut fra brukervennlighetshensyn, men det er likevel vanskelig å *dokumentere* en faktisk brukervennlighetsgevinst og slike systemer kommer relativt dårlig ut med hensyn til kapasitet. Derimot kan det argumenteres for identifikatorer som benytter bare et lite utvalg av symbolene i noen posisjoner, dersom disse posisjonene er tillagt et eksplisitt informasjonsinnhold, for eksempel de første seks posisjonene i dagens fødselsnummer.

Det å velge et alfabet med q symboler (for $q > 10$), blir derfor – i vår sammenheng – ensbetydende med å velge et alfabet som består av de ti desimalsifrene i mengden $\{0,1,2,3,4,5,6,7,8,9\}$, og $q-10$ bokstaver valgt fra mengden $\{\text{«A»}, \text{«B»}, \dots, \text{«Z»}\}$. Det finnes svært mange måter å velge slike mengder på. For $q=23$ finnes det for eksempel 10400600 måter å velge alfabetet på. Det er lett å argumentere for at mange av disse valgene vil være «dårlige», men vanskelig å påstå at ett bestemt alfabet er universelt bedre enn alle andre.

Dersom det har en betydelig negativ effekt at to bestemte symboler opptrer samtidig i det endelige alfabetet, kan vi beskrive dette som en *symbolkonflikt*. Vi kan gjøre en del rimelige men heuristiske antagelser for å definere en teoretisk modell for å kvantifisere effekten av slike symbolkonflikter. Et enkelt datamaskinprogram kan deretter identifisere et alfabet som beviselig, gitt modellen, er minst like godt som alle andre alfabet.

En slik modell kan utvikles på bakgrunn av flere momenter, som vi vil diskutere kort nedenfor.

- *Konvensjon og bakoverkompatibilitet.*

I dagens situasjon i Norge finnes der et eksisterende fødselsnummersystem som benytter et desimalt alfabet. Folk er vant til dette systemet, og i en ren utvidelse av systemet (som diskutert i vår tidligere rapport [1]) bør man velge samme alfabet (som i Alt.1) eller et nytt alfabet der det desimale alfabetet inngår som en del (Alt. 2.) Dersom man derimot ønsker en helt ny identifikator, for eksempel en *informasjonsløs* identifikator, kan det være en fordel å velge alfabetet slik at det går klart frem av format og alfabet hvilken type identifikator man har foran seg.

- *Robusthet*

Et riktig valg av system for kontrollsymbol(er) kan gi garantier for at systemet vil *detektere* visse registreringsfeil. Men et uheldig valg av *alfabet* kan øke sannsynligheten for at noen typer registreringsfeil *forekommer*. Uheldigvis er det mange forskjellige prosesser i registrering og behandling av identifikatorer som kan være involvert her, og disse prosessene kan være sårbar for feil på forskjellige måter og dermed sette forskjellige krav til alfabetvalg. Hovedregelen vil være at man ønsker å unngå *kritiske symbolkonflikter*, det vil si at man vil unngå at alfabetet inneholder par av symboler som «ligner på hverandre» og som derfor kan føre til feil ved registrering.

- *Lesing* (enten av menneskelig operatør eller gjennom automatisk tekstgjenkjenning) av skrevet/trykt identifikator: Her kan *grafisk likhet* skape problemer. For eksempel bør ikke bokstavene «I» og «O» forekomme som lovlige symboler i et alfabet som inneholder sifrene «1» og «0». Det finnes andre kombinasjoner av symboler som også er uheldige, og en full analyse er omfattende – samtidig som det er en overraskende mangel på tilgjengelige konkrete vitenskapelige forskningsresultater på dette temaet, selv i forhold til engelsk alfabet. Nedenfor vil vi se på eksempler som er relevante i vår situasjon.
- *Opplesing og «feilhøring»*: Noen par av symboler kan lettere enn andre føre til misforståelser ved muntlig overføring og ved «personlig uttale». Dette er imidlertid svært vanskelig å beskrive i en universell modell, og vi foretrekker å ikke prøve på det her.
- *Registrering med «datamaskin»-tastatur*: En vanlig årsak til generelle trykkfeil er at fingrene ligger feil på QWERTY-tastaturet, noe som fører til en type av såkalte *transkripsjonsfeil*. Vi antar at den mest vanlige feilen vil være at man taster inn horisontal nabotast til den korrekte tasten (for eksempel «S» istedenfor «A»), og at alle disse feilene er tilnærmet like sannsynlige. Vi ignorerer feiltypen der man taster vertikal nabotast til den korrekte tasten (for eksempel «Q» istedenfor «A») fordi disse feilene subjektivt (for oss) virker mindre sannsynlige. NB: Feil av denne typen kan i noen tilfelle utvides til doble transkripsjonsfeil, eller strekke seg over enda flere symboler. Tilsvarende må man forholde seg til *transposisjonsfeil*, en annen vanlig type av trykkfeil, der nabosymboler bytter plass. Feildeteksjonssystemene vi har foreslått vil i varierende grad ta hånd om disse typene av doble feil, og detektere de fleste eller alle av disse feilene.
- Registrering med Mobil/SMS vil potensielt innføre andre typer av registreringsfeil (for eksempel «B» istedenfor «A»). Det er uklart for oss om dette er en betydelig applikasjonstype i dag, og om det vil være det i fremtiden. For øyeblikket velger vi å ikke fokusere på feil som kan introduseres på denne måten.
- *Annen I/O-teknologi* (ikke «1/0-teknologi»): Maskinlesing, Strekkoder, RFID,...og annen fremtidig ny teknologi i et langt perspektiv. Dagens teknologi er svært forskjellig fra den som var tilgjengelig for femti år siden, og det er liten grunn til å tro at utviklingen vil bremse. Vi vil forvente at fremtidig teknologi vil være minst like robust som dagens, og kan ikke spekulere over dette i valg av modell.
- Husking og *feilhusking* er også prosesser som er involvert i feilmodeller, men er vanskelig å modellere (men se *brukervennlighet* nedenfor.) Vi velger imidlertid å ikke forsøke oss på å inkludere disse prosessene i vår enkle modell for symbolkonflikter.

Samlet sett stiller disse prosessene svært forskjellige krav til valg av alfabet. Dersom man skal ta hensyn til, og *unngå, alle* symbolkonflikter, vil man fort få problemer med å finne tilstrekkelig store alfabeter. En forenklet strategi kan gå ut på å ta hånd om de symbolkonflikter som man vurderer til å være «mest alvorlige», mens man ignorerer de

andre symbolkonfliktene og stoler på at kontrollsymbolene vil håndtere de fleste av de feil som oppstår.

- *Brukervennlighet* er vanskelig å definere. Det er et begrep som er basert på empiri i en populasjon med stadig skiftende digital kompetanse, og er vanskelig å måle. Brukervennlighet kan påvirke feilhusking. Forskning på *korttidsminne* indikerer at identifikatorer ikke bør være for lange (5-8 symboler). Dette kan være relevant dersom man, for eksempel i en saksbehandlerrolle, skal huske en annen persons identifikator over et kort tidsrom. I forhold til sin egen identifikator kan *langtidsminne* være mer relevant, og dersom personinformasjon av typen fødselsdato er inkludert vil man kunne huske tegnstrenger av større lengde. Vi kjenner ikke til forskningsresultater som relaterer alfabet og alfabetstørrelse til identifikatorlengde i forhold til korttids- og langtidsminne, og gjør også oppmerksom på at dette ligger langt utenfor vårt fomelle kompetanseområde.

Alfabet for hvert enkelt anbefalt alternativ.

Alternativ 1 og 3

Alternativene 1 og 3 er forutsatt med et desimalt alfabet, som i dagens fødselsnummer.

Alternativ 2

I vår opprinnelige rapport [1] er Alternativ 2 foreslått i forhold til tre forskjellige alfabetstørrelser, 17, 19 og 23. Alfabetstørrelsen 23 er mest robust i forhold til kapasitet og er vår hovedanbefaling, og er også brukt i Figur 1 i vår opprinnelige rapport [1]. Også for Alternativ 4 er vår anbefaling $q=23$ (og med seks informasjonssymboler.)

Ut fra diskusjonen om robusthet, kan vi spesifisere et alfabet basert på symbollikhet med hensyn på grafisk likhet, QWERTY-naboskap (og eventuelt audio-likhet og SMS-likhet, som vi ignorerer her). På grunn av manglende tilgjengelig forskning er vi nødt til å gjøre en del subjektive og heuristiske antagelser om hvor alvorlig hver symbolkonflikt er. (Vi antar at det er umulig å finne en teoretisk og universelt gyldig modell.) Når modellen først er valgt, kan et enkelt dataprogram produsere et optimalt alfabet (det vil si, et alfabet som *ut fra valgt modell* er beviselig minst like godt som alle andre alternative alfabet.)

Eksempel. Vi antar at alle ti desimalsifre skal være med i alfabetet, og at vi i tillegg skal velge et visst antall bokstaver fra A til Z. Anta at grafisk likhet mellom par av symboler (der minst ett ikke er et desimalsiffer) er gitt ved et tall fra 0 til 100. En enkel subjektiv modell for veiet symbolkonflikt på grunn av grafisk likhet kan beskrives ved tabellen nedenfor. I tabellen har vi tatt med også noen symbolkonflikter mellom par av desimalsifre («*interdesimale symbolkonflikter*»), selv om vi foreløpig ignorerer dem fordi vi uansett, for Alternativ 2, ønsker å beholde alle desimalsifre i det endelige alfabetet. Interdesimale symbolkonflikter er beskrevet i omtrentlig samsvar med empiriske data for desimalsifre fra Verhoeffs bok [3]. For hver symbolkonflikt (det vil si for hvert symbolpar som forekommer i tabellen) er det angitt en vekt som angir (subjektiv) grafisk likhet. Symbolpar som ikke forekommer i tabellen antas å ikke gi stor sannsynlighet for misforståelser.

Tabell 1: Grafisk likhet

0 - 6 : 50	1 - I : 100	4 - 7 : 20	6 - G : 100	C - O : 20	M - N : 10
0 - 8 : 50	1 - J : 80	4 - 9 : 40	7 - 9 : 30	D - O : 80	O - Q : 50
0 - 9 : 50	1 - L : 70	4 - A : 70	7 - I : 50	E - F : 10	P - R : 20
0 - D : 50	2 - 3 : 70	4 - I : 80	7 - T : 20	G - J : 20	Q - Y : 20
0 - O : 100	2 - 7 : 20	4 - L : 60	7 - Y : 20	G - Q : 60	U - V : 80
0 - Q : 80	2 - Z : 100	5 - 6 : 40	7 - Z : 20	G - Y : 50	V - W : 20
1 - 2 : 20	3 - 5 : 80	5 - S : 100	8 - B : 100	H - K : 20	

For QWERTY-likhet kan vi tilsvarende beskrive *horisontalt naboskap* mellom to nabotaster på tastaturet med en vekt på 10 (mens symbolpar som ikke er naboer på tastaturet får verdi 0.) I tillegg har vi tatt med konflikter mellom desimalsifre, som ofte kan tastes på to måter på tastaturet. For enkelhets skyld har vi representert dette ved en 50% sannsynlighet for bruk av øvre rad på QWERTY-tastaturet, og 50% sannsynlighet for bruk av eget numerisk tastatur (selv om denne forenklingen ikke nødvendigvis gjenspeiler virkeligheten.) Dette er vist i Tabell 2:

Tabell 2: Naboskap på tastatur

0 - 9 : 5	4 - 7 : 5	B - N : 10	G - H : 10	Q - W : 10
1 - 2 : 10	5 - 6 : 10	B - V : 10	H - J : 10	R - T : 10
1 - 4 : 5	5 - 8 : 5	C - V : 10	I - O : 10	T - Y : 10
2 - 3 : 10	6 - 7 : 5	D - F : 10	I - U : 10	U - Y : 10
2 - 5 : 5	6 - 9 : 5	D - S : 10	J - K : 10	X - W : 10
3 - 4 : 5	7 - 8 : 5	E - R : 10	K - L : 10	X - Z : 10
3 - 6 : 5	8 - 9 : 5	E - W : 10	M - N : 10	
4 - 5 : 10	A - S : 10	F - G : 10	O - P : 10	

Hvis vi summerer direkte vektene i de to tabellene, får vi Tabell 3:

Tabell 3: Samlet tabell for modell: Tabell 1 + Tabell 2

0 - 6 : 50	2 - 7 : 20	5 - 8 : 5	8 - B : 100	F - G : 10	O - P : 10
0 - 8 : 50	2 - Z : 100	5 - S : 100	9 - Q : 50	G - H : 10	O - Q : 50
0 - 9 : 55	3 - 4 : 5	6 - 7 : 5	A - S : 10	G - J : 20	P - R : 20
0 - D : 50	3 - 5 : 80	6 - 8 : 30	B - H : 30	G - Q : 60	Q - W : 10
0 - O : 100	3 - 6 : 5	6 - 9 : 5	B - N : 10	G - Y : 50	Q - Y : 20
0 - Q : 80	3 - 8 : 40	6 - C : 30	B - V : 10	H - J : 10	R - T : 10
1 - 2 : 30	3 - B : 20	6 - G : 100	C - O : 20	H - K : 20	T - Y : 10
1 - 4 : 35	4 - 5 : 10	7 - 8 : 5	C - V : 10	I - L : 30	U - V : 80
1 - 7 : 20	4 - 7 : 25	7 - 9 : 30	D - F : 10	I - O : 10	U - Y : 10
1 - I : 100	4 - 9 : 40	7 - I : 50	D - O : 80	I - U : 10	V - W : 20
1 - J : 80	4 - A : 70	7 - T : 20	D - S : 10	J - K : 10	X - W : 10
1 - L : 70	4 - I : 80	7 - Y : 20	E - F : 10	J - Y : 30	X - Z : 10
2 - 3 : 80	4 - L : 60	7 - Z : 20	E - R : 10	K - L : 10	

Både hver av de to enkelttabellene og måten de er kombinert på kan diskuteres, og man kan også innvende at andre symbolkonflikter bør ta hensyn til. Vi vil likevel hevde at eksempelmodellen gir et ikke ufornuftig utgangspunkt for å velge alfabet.

Et lite dataprogram kan nå finne det alfabetet der summen av verdiene av gjenværende symbolkonflikter er lavest mulig. Dette er en forenklet heuristikk; for eksempel kan man også ha et ønske om å minimere *høyeste* vekt blant gjenværende symbolkonflikter. Vi hopper over detaljene, og ettersom tabellene uansett er heuristisk konstruert, velger vi likevel denne forenklete modellen.

Gitt modellen finnes det ikke bedre alfabet enn de som er gitt i Tabell 4, for forskjellige relevante verdier av q .

Tabell 4: Beste alfabet for modellen gitt ved Tabell 3, dersom alle desimalsifre skal være med

q : 17	Max	Sum
0 1 2 3 4 5 6 7 8 9 F K N R U W X	210	1370
q : 19		
0 1 2 3 4 5 6 7 8 9 F K M N P T U W X	210	1450
q : 23		
0 1 2 3 4 5 6 7 8 9 C E F H K M N P T U W X Y	210	1670
q : 29		
0 1 2 3 4 5 6 7 8 9 A C D E F H K M N P R S T U V W X Y Z	250	2730

I Tabell 4 har vi gitt alfabetene for $q = 17, 19, 23$ og 29 , som er aktuelle primtallsverdier. For hvert alfabet har vi angitt en «Max»-verdi og en «Sum»-verdi. Disse er beregnet slik:

- For hvert gjenværende symbol i alfabetet beregner vi summen av verdier av symbolkonflikter («konflikt-verdi») som involverer dette symbolet og et av de andre gjenværende symbolene i alfabetet. Dersom et symbol får en høy konfliktverdi, indikerer dette at dette symbolet har en (relativt) høy sannsynlighet for å bli involvert i en symbolfeil. «Max» i tabellen representerer den høyeste «konflikt-verdien» blant symbolene i alfabetet.
- «Sum» i tabellen representerer summen av «konflikt-verdier» for alle symboler i alfabetet. Merk at dette tallet er to ganger summen av verdier for alle gjenværende symbolkonflikter i alfabetet.

Dersom vi tillegger tastaturfeil dobbelt så mye vekt, får vi den alternative modellen:

Tabell 5: Samlet tabell for alternativ modell: Tabell 1 + 2* Tabell 2

0 - 6 : 50	2 - 7 : 20	5 - 8 : 10	8 - B : 100	F - G : 20	O - P : 20
0 - 8 : 50	2 - Z : 100	5 - S : 100	9 - Q : 50	G - H : 20	O - Q : 50
0 - 9 : 60	3 - 4 : 10	6 - 7 : 10	A - S : 20	G - J : 20	P - R : 20
0 - D : 50	3 - 5 : 80	6 - 8 : 30	B - H : 30	G - Q : 60	Q - W : 20
0 - O : 100	3 - 6 : 10	6 - 9 : 10	B - N : 20	G - Y : 50	Q - Y : 20
0 - Q : 80	3 - 8 : 40	6 - C : 30	B - V : 20	H - J : 20	R - T : 20
1 - 2 : 40	3 - B : 20	6 - G : 100	C - O : 20	H - K : 20	T - Y : 20
1 - 4 : 40	4 - 5 : 20	7 - 8 : 10	C - V : 20	I - L : 30	U - V : 80
1 - 7 : 20	4 - 7 : 30	7 - 9 : 30	D - F : 20	I - O : 20	U - Y : 20
1 - I : 100	4 - 9 : 40	7 - I : 50	D - O : 80	I - U : 20	V - W : 20
1 - J : 80	4 - A : 70	7 - T : 20	D - S : 20	J - K : 20	X - W : 20
1 - L : 70	4 - I : 80	7 - Y : 20	E - F : 10	J - Y : 30	X - Z : 20
2 - 3 : 90	4 - L : 60	7 - Z : 20	E - R : 20	K - L : 20	

Blant de beste alfabetene for modellen i Tabell 5 finner vi de som er listet opp i Tabell 6:

Tabell 6: Beste alfabet for modellen gitt ved Tabell 5, dersom alle desimalsifre skal være med

q : 17	Max	Sum
0 1 2 3 4 5 6 7 8 9 F K N R U W X	230	1560
q : 19		
0 1 2 3 4 5 6 7 8 9 F H K N P T U W X	230	1600
q : 23		
0 1 2 3 4 5 6 7 8 9 C E F H K M N P T U W X Y	230	1940
q : 29		
0 1 2 3 4 5 6 7 8 9 A C D E F H K M N P R S T U V W X Y Z	280	3140

Vi ser at de anbefalte alfabetene er nesten de samme som, og i noen tilfeller helt like, de i forrige modell. Det kan være en indikasjon på at modellen er robust for mindre endringer i tabellen.

Legg merke til at «Sum»- og «Max»-verdiene er litt høyere for Tabell 6 enn for Tabell 4. Dette skyldes at verdiene i Tabell 5 er lik eller høyere den tilsvarende verdien i Tabell 3 for hver symbolkonflikt.

Alternativ 4

Også for Alternativ 4 er vår hovedanbefaling $q=23$ (og med seks informasjonssymboler.) Her er det ikke opplagt at alle desimalsifre bør være med. Vi finner alfabetene gitt i Tabell 7.

Tabell 7: Beste alfabet for de to modellene gitt ved Tabell 3, dersom ikke alle desimalsifre trenger å være med

q : 17	Max	Sum
1 5 9 A B C D E G K M P T U W X Z	10	40
q : 19		
2 5 8 9 A C D E F H J L M N P T U W X	20	150
q : 23		
2 5 8 9 A C D E F G H I J K L M N P R T U W X	40	470
q : 29		
3 6 7 9 A B C D E F H I J K L M N O P Q R S T U V W X Y Z	170	1750

I forhold til alfabetene i Tabell 4 (for bruk i Alternativ 2 og modellen beskrevet i Tabell 3, er alfabetene i Tabell 7 adskillig mer robust, som vi ser ved «Max»- og «Sum»-verdiene (som er direkte sammenlignbare for Tabell 4 og Tabell 7, fordi de begge er basert på Tabell 3.) Det vil si, *dersom modellen er rimelig*, er sannsynligheten for feil ved tegn-for-tegn karaktergjenkjenning for feildeteksjon en god del lavere for Alternativ 4 enn for Alternativ 2. Faktisk utgjøres mesteparten av de alvorlige symbolkonfliktene i alfabetene i Tabell 6 av innbyrdes konflikter mellom desimalsymboler. To eksempler:

- For alfabetet for $q=17$ i Tabell 4 er *alle* gjenværende symbolkonflikter av typen «interdesimal». Ved bruk av dette alfabetet kan man forvente en lavere *gjennomsnittlig* feilsannsynlighet enn for et rent desimalt alfabet.

- For $q=23$ vil sannsynligheten for karaktergjenkjenningsfeil ved alfabetet gitt i Tabell 4 ikke være mye høyere enn den tilsvarende sannsynligheten for feil dersom alfabetet er desimalt. Feilsannsynligheten for enkeltkaraktergjenkjenning er vesentlig lavere for det «store» alfabetet i Tabell 7 med 23 elementer enn det er for et desimalt alfabet. Dette er en effekt som kommer *i tillegg til* de forskjeller i robusthet gitt ved feildeteksjons-egenskaper som vi har diskutert i Figur 1 i vår opprinnelige rapport [1].

Vi bemerker forøvrig at automatisk optisk karaktergjenkjenning (OCR) vanligvis fungerer bedre når karaktergjenkjenning blir gjennomført i en *kontekst*, det vil si at istedenfor å forsøke å bare gjenkjenne symboler tegn for tegn vil OCR-programmet prøve å tolke grafiske mønstre i forhold til hvordan de enkelte symboler passer inn i en tekstlig sammenheng. En slik kontekst kan være vanskeligere å definere når man bare har en kort identifikatorstreng. Men dersom man skal utføre OCR på data med identifikatorer (og man vet at det er det som skal leses) er det mulig å bruke kontrollsymbolene til å skape en slik tekstlig sammenheng, og dette kan forbedre presisjonen i karaktergjenkjenningen kraftig. Dette vil fungere bedre med Alternativ 2 og 4 enn med de andre alternativene fordi feilkorreksjonsegenskapene er mye bedre. Vi understreker at dette er en alternativ bruk av kontrollsymbolene.

Konklusjon: En pseudoteoretisk tilnærming tilsier at valg av alfabet kan ha betydelig innflytelse på sannsynlighet for feil enkeltsymbolgjenkjenning, og som konsekvens også på sannsynlighet for uoppdaget feilregistrering. Generelt vil en ren *utvidelse* av alfabetet aldri redusere sannsynligheten for feil enkeltsymbolgjenkjenning. For eksempel vil alle alfabeter som inneholder alle desimalsiffer (og noen symboler til), automatisk inneholde alle «interdesimale» symbolkonflikter, selv om «noen symboler til» i mange tilfeller kan velges slik at de alvorligste symbolkonfliktene er de interdesimale og den *gjennomsnittlige* sannsynligheten for enkeltsymbolgjenkjenning kan reduseres i forhold til et rent desimalt alfabet. Dersom vi kan velge å ikke ha med alle desimalsymboler i alfabetet kan vi forbedre robustheten kraftig. Vi vil igjen understreke at modellene vi har brukt i Tabell 3 og 5 er relativt enkle og at de nok bør utredes mer grundig, både teoretisk og empirisk, før et endelig valg foretas.

3) Kompatibilitet og sameksistens med dagens identifikator

Identifikatorsystemene vi har foreslått i vår tidligere rapport [1] er ment enten å fungere parallelt med dagens fødselsnummersystem, eller helt uavhengig av det.

Nedenfor beskriver vi dette i nærmere detalj for hvert alternativ.

Informasjonsbærende identifikatorer

Ved bruk av en identifikator som inneholder informasjon kan og bør de som har fått et fødselsnummer etter gammel ordning få beholde dette. Nye numre kan håndteres på følgende måte.

Informasjonsbærende identifikatorer: Tildeling

Vi ser for oss at tildeling av nye nummerserier foregår sentralisert, og at det ikke tilkommer nye momenter i forhold til dagens ordning. Forutsatt robust datanettforbindelse kan tildeling foregå umiddelbart, for eksempel i det øyeblikk en fødsel er registrert. Det vil redusere behov for såkalte helsenumre.

Informasjonsbærende identifikatorer: Verifikasjon

Ved sameksistens mellom flere typer identifikator må et verifiseringsprogram sjekke et oppgitt fødselsnummer mot flere verifikasjonsalgoritmer. Se detaljer nedenfor for Alternativene 1 og 2.

Informasjonsbærende identifikatorer: Oppslag i databaser

Når identifikatoren er verifisert, ser vi ikke for oss at det skal innebære noe problem å bruke denne som et vilkårlig oppslagsnøkkel mot en database. Vi tar forbehold om at databasen i seg selv foretar en egen verifisering.

Alternativ 1 og dagens fødselsnummersystem

I dette alternativet vil de som har fødselsnummer i dagens ordning beholde dette. Når dagens ordning ikke har flere ledige identifikatorer vil man skifte automatisk til nytt system i Alternativ 1 for nye brukere. Verifikasjon: En datamaskin kan sjekke kontrollsymboler og utføre feilkontroll for både Alternativ 0 (dagens fødselsnummer) og Alternativ 1. En gitt lovlig identifikator vil tilfredsstille nøyaktig én av disse testene, og feilkontrollen vil også (se Figur 1 i [1]) beskytte mot feil som eventuelt kunne forandre en Alternativ 0-identifikator til å ligne på en Alternativ 1-identifikator og omvendt. Dette er transparent for brukerne: Brukere behøver ikke vite (eller tenke over) hvilken type av fødselsnummer de har fått tildelt.

Alternativ 2 og dagens fødselsnummersystem

I dette alternativet vil de som har fødselsnummer i dagens ordning beholde dette. Når dagens ordning ikke har flere ledige identifikatorer vil man skifte automatisk til nytt system i Alternativ 2 for nye brukere. Verifikasjon: En datamaskin kan sjekke kontrollsymboler og utføre feilkontroll for både Alternativ 0 (dagens fødselsnummer) og Alternativ 2. En gitt lovlig identifikator vil tilfredsstille nøyaktig én av disse testene, og feilkontrollen vil også (se Figur 1 i [1]) beskytte mot feil som eventuelt kunne forandre en Alternativ 0-identifikator til å ligne på en Alternativ 2-identifikator og omvendt. Feilkontrollen vil være mer effektiv og robust enn for Alternativ 1. En bruker vil kunne se at fødselsnummeret er av ny type fordi «løpenummeret» har minst tre bokstaver.

Informasjonsløse identifikatorer

Dersom en informasjonsløs identifikator foretrekkes, antar vi at det er ut fra et politisk ønske om å unngå informasjonslekkasje. Det samme politiske ønsket kan tilsi at alle eksisterende brukere får nye identifikatorer. Imidlertid vil vi anta at det uansett vil være nødvendig å håndtere gamle identifikatorer i lang tid fremover, på grunn av en eksisterende base av gamle data indeksert med gamle fødselsnumre.

Informasjonsløse identifikatorer: Tildeling

Vi ser for oss at tildeling av nye nummerserier foregår sentralisert, og at det ikke tilkommer nye momenter i forhold til dagens ordning. Forutsatt robust datanettforbindelse kan tildeling foregå umiddelbart, for eksempel i det øyeblikk en fødsel er registrert. Det vil redusere behov for såkalte helsenumre.

Informasjonsløse identifikatorer: Verifikasjon

Ved sameksistens mellom flere typer identifikator må et verifiseringsprogram sjekke et oppgitt fødselsnummer mot flere verifikasjonsalgoritmer. Dette gjelder både for Alternativene 3 og 4, og vi diskuterer ikke detaljene nærmere.

Informasjonsløse identifikatorer: Oppslag i databaser

Når identifikatoren er verifisert, ser vi ikke for oss at det skal innebære noe problem å bruke denne som et vilkårlig oppslagsnøkkel mot en database. Vi tar forbehold om at databasen i seg selv foretar en egen verifisering.

Alternativ 3

Alternativ 3 er en informasjonsløs identifikator. Uavhengig av generelle betraktninger som nevnt ovenfor, gjelder det i prinsippet at også for dette alternativet kan de som er tildelt fødselsnummer i dagens ordning beholde dette.

Rommet, eller mengden, av mulige identifikatorer overlapper med fødselsnummer- og D-nummerseriene. Det er for såvidt en stor reserve av identifikatorer tilgjengelig (det vil si, ikke brukt i fødselsnummer- eller D-nummerserier) innenfor dette rommet, men de ligger «plundrete» til og det kan skape forvirring dersom man bruker en tidligere ubrukt identifikator som for eksempel svarer til en bestemt dato på DDMMÅÅ-format i de første seks sifrene. Med andre ord må mange nummerserier forkastes. I skrivende stund er det ikke klart hvilke nummerserier dette gjelder. En «renslig» bruk av Alternativ 3 kan redusere mengden av tilgjengelige identifikatorer ned mot et uforsvarlig lavt nivå, avhengig av hvilke nummerserier som forkastes. I neste rapport kommer vi tilbake til dette, ut fra mer detaljerte forutsetninger for hvilke nummerserier som forkastes.

Alternativ 4

Alternativ 4 er en informasjonsløs identifikator. Denne har et vesentlig forskjellig format i forhold til dagens fødselsnummer, og dette tilsier at også alle eksisterende brukere bør få tildelt en ny identifikator. Formatet er så forskjellig (forskjellig alfabet i forhold til dagens fødselsnummer og 7/8 tegn langt i forhold til 11) at det er svært lett å se forskjell på Alternativ 4 og dagens fødselsnummer.

Konklusjon: For de av de vurderte identifikatorene der det er aktuelt, er sameksistens med dagens fødselsnummer ikke noe teknisk problem.

4) Kommentar vedrørende Alternativ 7

Vi har fått oversendt e-postkorrespondanse mellom Boris Schürmann og K. T. Homme i forbindelse med et konkret forslag:

«Gitt disse enkle premissane kunne ein rett og slett fristille dei to fyrste siffera i dagens fødselsnummer, og avskaffe skiljet mellom partal og oddetal for kjønn.»

Såvidt vi kan se er forslaget bortimot identisk med Alternativ 7, som vi har diskutert tidligere i vår opprinnelige rapport [1], og lagt til side fordi vi anser at dette forslaget er mindre gunstig enn de anbefalte alternativ.

Referanser

[1] Tor Helleseeth og Øyvind Ytrehus, « Rapport om egenskaper ved ny personidentifikator», *Rapport for Skattedirektoratet*, 23.04.2012.

[2] Grunndatarapporten,
http://www.regjeringen.no/nb/dep/fin/dok/rapporter_planer/rapporter/2007/utveksling-av-grunndata-pa-personinforma.html?id=491849

[3] J. Verhoeff, “Error Detecting Decimal Codes”, Mathematical Centre Tract 29, The Mathematical Centre, Amsterdam, 1969.

Rapport om D-nummer og økt populasjon for personidentifikator
Tor Helleseeth og Øyvind Ytrehus
Universitetet i Bergen/Selmersenteret
Revidert 04.04.2013

Innhold

Innledning.....	2
1) Kapasitet for D-nummer	4
2) Robusthet for D-nummer.....	7
2.1 Status 2012	7
2.2 Beregning av fremtidig behov for D-numre i forhold til kapasitet i vid forstand	9
2.3 Kapasitet i vid forstand: Resultater.....	11
2.4 Kapasitet i snever forstand.....	13
3) Robusthet for fødselsnummer ved økt populasjon/personkrets	16
Kapasitet for Alternativ 1 og Alternativ 2	17
Kapasitet i vid forstand : D+F-nummer for Alternativ 1 og Alternativ 2	17
Kapasitet i snever forstand : D+F-nummer for Alternativ 1 og Alternativ 2.....	19
Informasjonsløse identifikatorer	22
4) Øke kapasitet for D-nummer.....	23
5) Kapasitet for fødselsnummer alternativ 3	24
6) Andre kommentarer	26
Appendiks A: Vekst i populasjoner: Scenarier.....	27
D-nummer-populasjon	27
Fødselsnummer	28
Kombinert populasjon bestående av dagens fødselsnummer- og D-nummerpopulasjoner..	29
Fødselsnummer, D-nummer og DUF-nummer	30
Helsennummer	33
Appendiks B: Aldersfordeling ved tildeling av identifikator	34
Referanser	38

Innledning

Denne rapporten er en fortsettelse av [1] og [2], som vi forutsetter lest og kjent. I denne rapporten besvarer vi en rekke konkrete spørsmål som Skattedirektoratet har stilt oss [3] i forbindelse med en utredning om kapasitet og behov for nye ordninger for personidentifikatorer. Vi henviser til de enkelte kapitlenes konklusjoner for hver enkelt problemstilling.

Innledningsvis ønsker vi å klargjøre en del begreper.

Kapasitet: Begrepet *kapasitet* uttrykker, innenfor konteksten av denne rapporten, en numerisk beskrivelse av hvor mange forskjellige identifikatorverdier som kan tildeles for et gitt valg av identifikatorsystem. De forskjellige identifikatorsystemene er forskjellige i bruk, og det er derfor hensiktsmessig å bruke forskjellige eksakte definisjoner i hvert tilfelle.

- For en *informasjonsløs* identifikator er det i prinsippet ingen individinformasjon innbakt i identifikatorverdien tilordnet hvert individ. Identifikatorverdier kan derfor tildeles fortløpende ut fra den totale mengden av «lovlige» identifikatorverdier. Dette gir en effektiv utnyttelse av mengden av lovlig verdier, og det gir mening å definere kapasiteten ganske enkelt som det totale antallet forskjellige identifikatorverdier.
- De *informasjonsbærende* identifikatorer som er i bruk, som D-nummer og fødselsnummer, og også de som har vært vurdert i forbindelse med fremtidige løsninger, inneholder følgende individinformasjon: fødselsdag, måned, og århundre, samt kjønn. Her kan vi vurdere kapasitet i to forskjellige perspektiv:
 - Kapasitet i *snever* forstand: Det er ønskelig at individinformasjonen er korrekt for *hvert enkelt* individ. Dette forutsetter at antall tilgjengelige identifikatorverdier for hver enkelt dag og for hvert kjønn er tilstrekkelig stort til å dekke behovet. Det er derfor fornuftig å definere **kapasitet i snever forstand** som **antall tilgjengelige identifikatorverdier pr. kjønn og dag**. Kapasiteten i snever forstand tilfredsstiller behovet bare dersom det i populasjonen, for alle mulige fødselsdager og for hvert kjønn, ikke finnes flere individer enn denne kapasiteten.
 - Kapasitet i *vid* forstand: En kapasitet i snever forstand gir mening i en populasjon der antall individer fordeler seg (tilfeldig men) noenlunde jevnt over alle årets dager og over begge kjønn. Dette har tradisjonelt vært oppfylt for fødselsnumre. I den senere tid utgjør innvandring en stor del av nye tildelinger av fødselsnumre, og en konsekvens av dette er at populasjonen i mindre grad fordeler seg jevnt over alle datoer og begge kjønn. Det samme gjelder populasjonen som blir tildelt D-numre. Typisk vil behovet for identifikatorverdier være betydelig høyere for menn enn for kvinner, og mye høyere for datoer som 1. januar, 1. juli, 31. desember, på grunn av at populasjonen inneholder personer fra kulturer som ikke har tradisjon for eksakt registrering av fødselsdato.

Dersom kapasiteten i snever forstand i en slik populasjon skal oppfylle behovet må man enten tilordne mange flere identifikatorverdier til de kritiske datoene (det er teknisk mulig å velge en koding som oppfyller dette kravet, men et ønske om «humanlesbarhet» tilsier at slike alternativ ikke har vært vurdert seriøst), eller sørge for at *alle* datoer og begge kjønn må ha «stor nok» kapasitet. Den pragmatiske løsningen på dette problemet er at når et nytt individ skal tildeles en identifikator og det ikke finnes flere identifikatorer som samsvarer med individets kjønn og offisielle fødselsdato, blir det valgt en identifikatorverdi for annen «vedtaksdato» innenfor samme år (og for samme kjønn).

Denne pragmatiske saksbehandlingen er mulig hvis den totale tilveksten i

populasjonen ikke overstiger den totale tilgangen på identifikatorverdier pr. år. Det gir derfor mening å definere **kapasitet i vid forstand** for en informasjonsbærende identifikator som **antall identifikatorverdier pr.år pr. kjønn**. Dersom tilveksten i populasjonen pr. år overstiger denne vid-forstand-kapasiteten er vedtaksdatometoden ikke lenger mulig.

Merk at kapasitet, for hver definisjon, er deterministisk og veldefinert matematisk, og at de numeriske verdiene vi oppgir i dette dokumentet er eksakte (innenfor den presisjonen som er hensiktsmessig.)

Behov. I motsetning til kapasiteten for et gitt identifikatorsystem, er behovet for identifikatorer i høyeste grad stokastisk og svært vanskelig å forutse med en rimelig grad av sikkerhet. Behovet vil avhenge av

- Politiske forhold,
- Økonomiske forhold, særlig Norge i forhold til omverdenen,
- Politiske beslutninger om administrative rutiner:
 - Hvilken populasjon skal omfattes av en gitt identifikator?
 - Velferdsordninger, som for eksempel vil påvirke tildeling av D-nummer for barn,
- Kriger,
- Epidemier,
- Naturkatastrofer,
- ...og helt sikkert mange andre ting.

På grunn av disse forhold er det ikke mulig for oss å kvantifisere sannsynligheter i forhold til robusthet av identifikatorsystem. Vi vil istedenfor forholde oss til et utvalg av **scenarier**, diskutert i Appendiks A, og til en viss grad argumentere for vårt syn på hvorvidt de enkelte scenariene er fornuftige.

Resten av denne rapporten er strukturert som følger: Hvert kapittel adresserer én av oppgavene i forespørselen. I to appendiks inkluderer vi beskrivelser av modeller og verktøy vi har brukt.

1) Kapasitet for D-nummer

Dette kapittelet er et svar på følgende oppgave fra [3]:

Oppgave 1

[Tabellen] er basert på opplysninger fra Folkeregisteret om hvilke individnummer som er brukt/brukes for personer født fra 1858 og potensielt fram til 2057. Det bes om at oppstillingen og beregnet kapasitet for de ulike tidsperiodene gjennomgås og kvalitetssikres.

Praksis for tildeling av D-nummer for personer født i forskjellige århundre følger Tabell 1:

Første siffer i individnummer	Født i perioden	Individnummer som kan brukes	Kapasitet pr. dag
5	1858-1899	500-599	82
0,1,2,3,4	1900-1939	000-499	413
0,1,2,3,4,(9 reserve)	1940-1999	000-499 (900-999)	413 (495)
5,6,7,8,9	2000-2039	500-999	413
5,6,7,8	2040-2057	500-899	330
9 (reserve, hvis 9 ikke brukes 1940-1999)	2040-2057	900-999	82

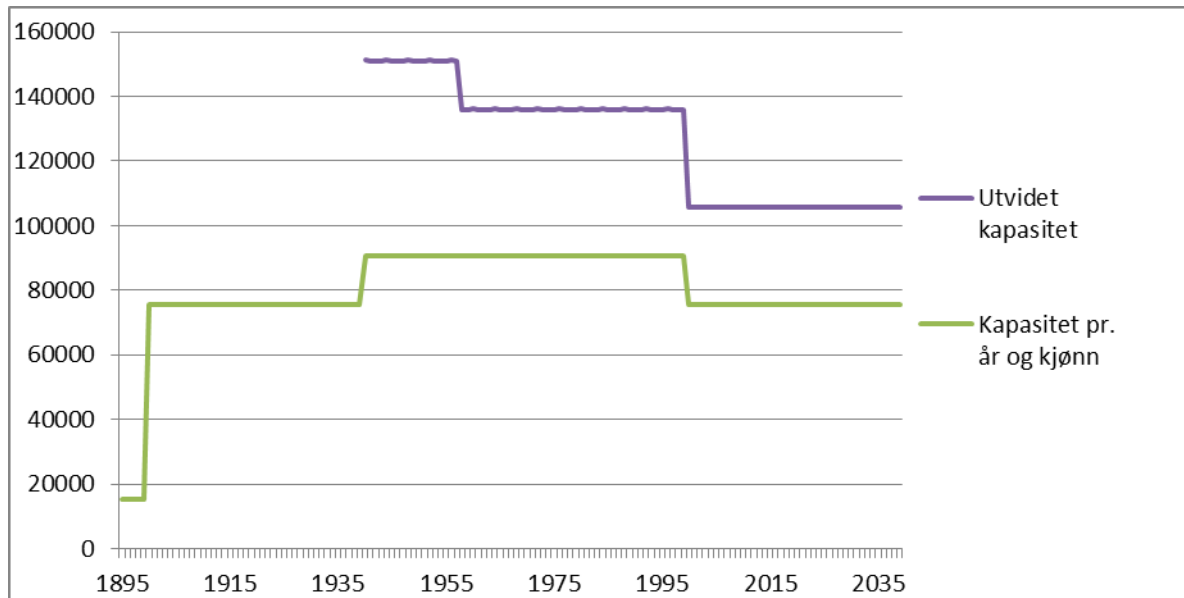
Tabell 1: Sammenhengen mellom det første siffer i individnummer, fødselsår, og kapasiteten til D-nummeret. Som for fødselsnummeret blir ikke alle individnumre godkjent pga. kontrollberegningen. Kolonnen «Kapasitet pr. dag» gjelder **sum for begge kjønn**. Alle tall i denne kolonnen er rundet ned til nærmeste heltall.

Tabell 2 viser en alternativ måte å fremstille bruken av fødselsnummerserier.

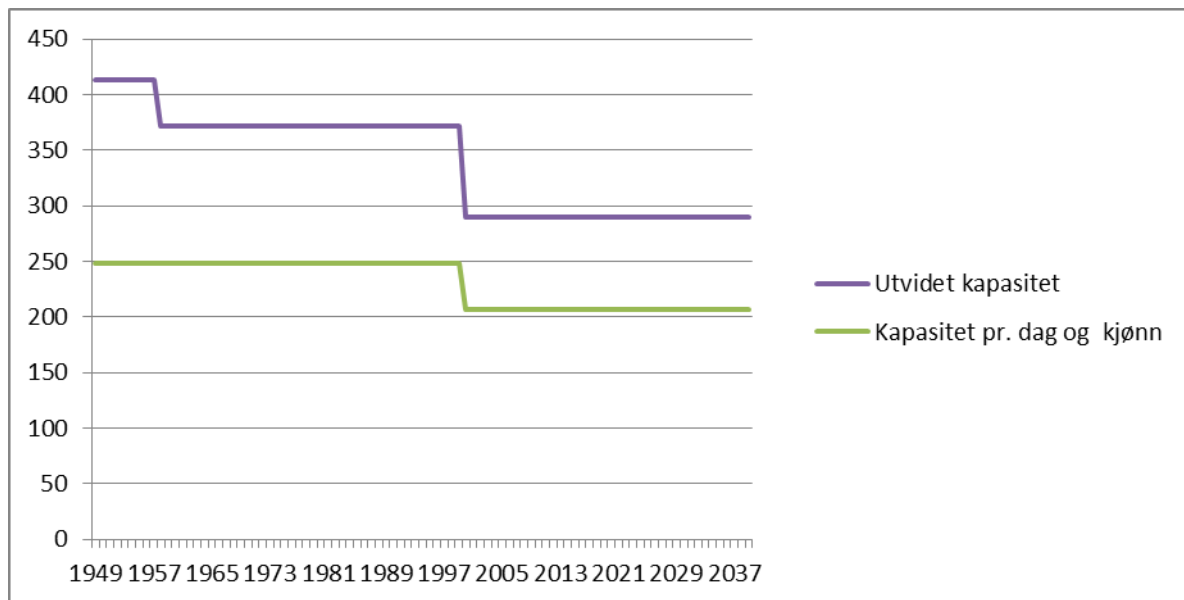
Første siffer ↓	D-Nummer: To siste siffer i årstall		
	00-39	40-57	58-99
0	1900-39	1940-1957	1958-1999
1			
2			
3			
4			
5	2000-2039	2040-2057	1858-1899
6			Ikke i bruk
7			
8			1958-1999 (Reserve)
9			
		1940-1957 (Reserve)	
		2040-2057 (Hvis ikke brukt)	Ikke i bruk

Tabell 2 Koding av århundre. For hvert første siffer er det i gjennomsnitt 82,6 gyldige løpenumre pr fødselsdato.

Kapasiteten i henholdsvis vid og snever forstand er vist i Figur 1 og Figur 2.



Figur 1 Kapasitet i vid forstand, det vil si pr kjønn pr. år, frem til 2040. «Utvidet kapasitet» refererer til forslag nedenfor under «Kommentarer».



Figur 2 Kapasitet i snever forstand, det vil si pr kjønn pr. dag. «Utvidet kapasitet» refererer til forslag nedenfor under «Kommentarer».

Kommentarer

- På grunn av tildeling av **5-serien** til personert født 1858-1899, er denne tabellen **ikke** i samsvar med **Forskrift om folkeregistrering fra 2007, § 2-5**, som sier at «*Som individsiffer for personer født:*
 - *før 1. januar 2000 brukes serien 000-499,*
 - *etter 1. januar 2000 brukes serien 500-999.....»*
- I Figur 1 og Figur 2 har vi allerede antatt at reserveserien (9-serien) tas i bruk for fødselsårene 1940-1957 og for 1948-1999. Kapasiteten (i snever forstand) pr. dag vil da være 495 pr. dag for disse fødselsårgruppene, eller 247 pr. dag pr. kjønn.
- For 1958-99 **kan 6,7,8 serien tas i bruk**, ettersom D-nummerserien ikke er ment for bruk etter 2057. Dette øker kapasiteten til 743 pr. dag, eller 371 pr. kjønn pr. dag.
- For 1940-57, **kan 5,6,7,8-serien tas i bruk**, dersom det finnes en annen løsning for årene etter 2040. (Analysen i kapittel 2 indikerer at dette vil bli nødvendig.) Kapasitet øker da til maksimalt 826 pr. dag, eller 413 pr. kjønn pr. dag.
- For 2000-2039 vil kapasiteten i utgangspunktet være begrenset til 413 per dag, ettersom 5 nummerserier er tildelt fødselsårgruppen 1900-1939. På grunn av at de fleste D-nummer i denne gruppen blir tildelt voksne individ, vil eventuelle problemer uansett neppe oppstå før etter 2020, hvis vi tar utgangspunkt i aldersfordeling for de som allerede har fått D-nummer [4].
- Selv om årsgruppen 1900-39 er tildelt 5 nummerserier (0-4), ser vi av statistikken [4] at faktisk bruk i antall er mye mindre. Dersom nummerserier kan reallokeres, kan kapasiteten for 2000-2039 økes betydelig. Dette kommer også an på i hvilken grad fødselsdatoer er fordelt i 1900-1939-gruppen. **I Figur 1 og 2 har vi antatt at 2 nummerserier kan reallokeres til 2000-2039.**

2) Robusthet for D-nummer

Dette kapittelet er et svar på følgende oppgave fra [3]:

Oppgave 2

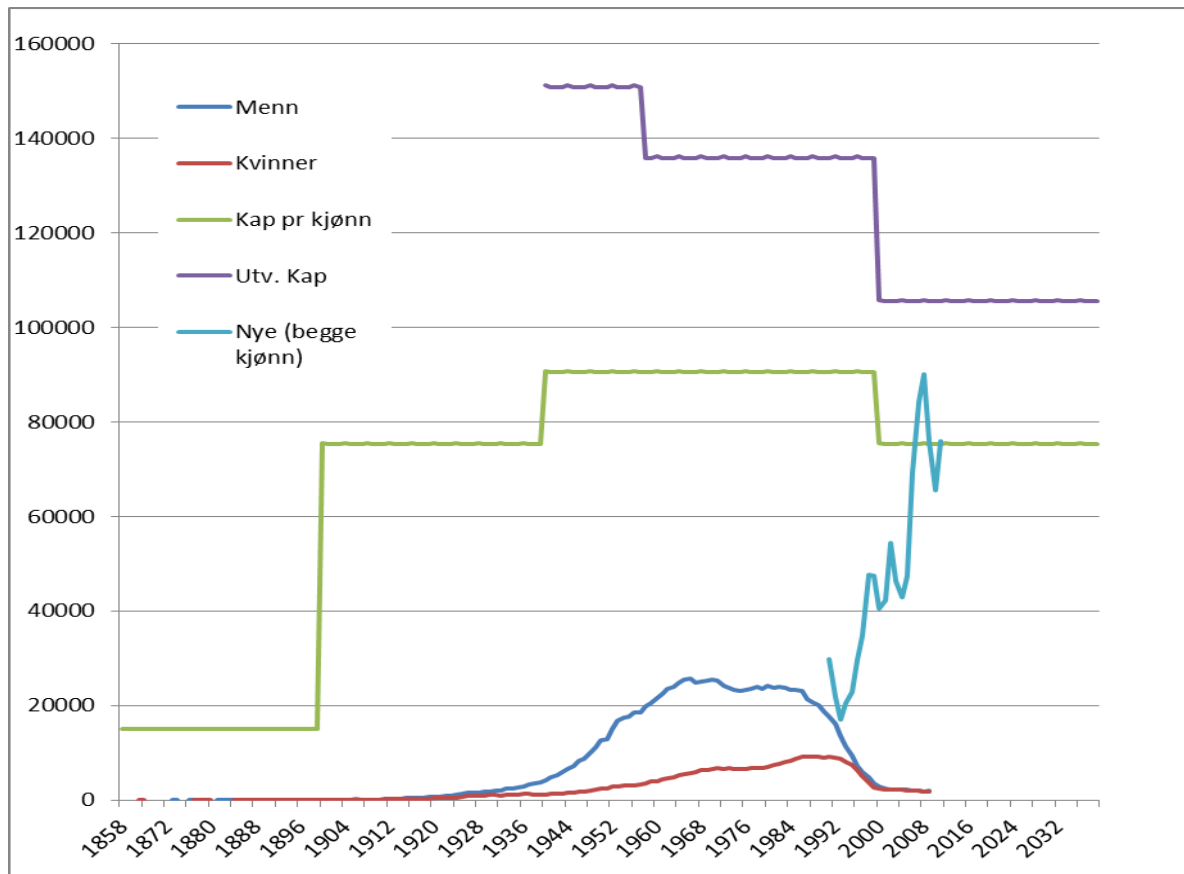
Oppgaven er å beregne ca. hvilket år kapasiteten til dagens D-nummer med stor sannsynlighet vil bli sprengt for flere datoer, med utgangspunkt i de tre vekstscenariene <gitt i [3]>. Beregningen gjennomføres fram til år 2150.

Vi vil understreke at analysen er gjort **i henhold til de spesifikke vekstscenariene**. Videre gjør vi en antagelse med hensyn på **alders- og kjønnsfordeling** i D-nummerpopulasjonen ved tildelingstidspunktet (se Appendiks B.) Vi tar ikke stilling til sannsynlighet for de forskjellige scenariene.

Hovedkonklusjon: D-nummerordningen har tilstrekkelig kapasitet frem til ca 2030 under alle de vurderte scenarier. Fra og med 2040 vil det uansett være nødvendig å etablere et nytt system, med mindre veksten i D-nummerpopulasjonen viser seg å bli svært lav.

2.1 Status 2012

Figur 3 viser status for populasjonen som er tildelt D-nummer ved utgangen av 2011. De blå og røde kurvene viser fordeling av fødselsår for henholdsvis menn og kvinner. For menn er foreløpig 1966 det fødselsåret som har flest D-nummertildelinger (25664); blant kvinner som har fått D-nummer er 1989 det vanligste fødselsåret (9276). Videre har vi tegnet inn kapasitet i vid forstand og utvidet kapasitet i vid forstand, som diskutert i Kapittel 1. Vi ser at behovet er vesentlig mindre enn kapasitet i vid forstand. Endelig, for å se disse tallene i sammenheng med utviklingen, har vi tegnet inn antall D-numre tildelt (sum for begge kjønn) for perioden 1991-2011. Vi ser at denne kurven øker kraftig i de senere årene, sammenlignet med størrelsen på årskullene i eksisterende D-nummer-populasjon.

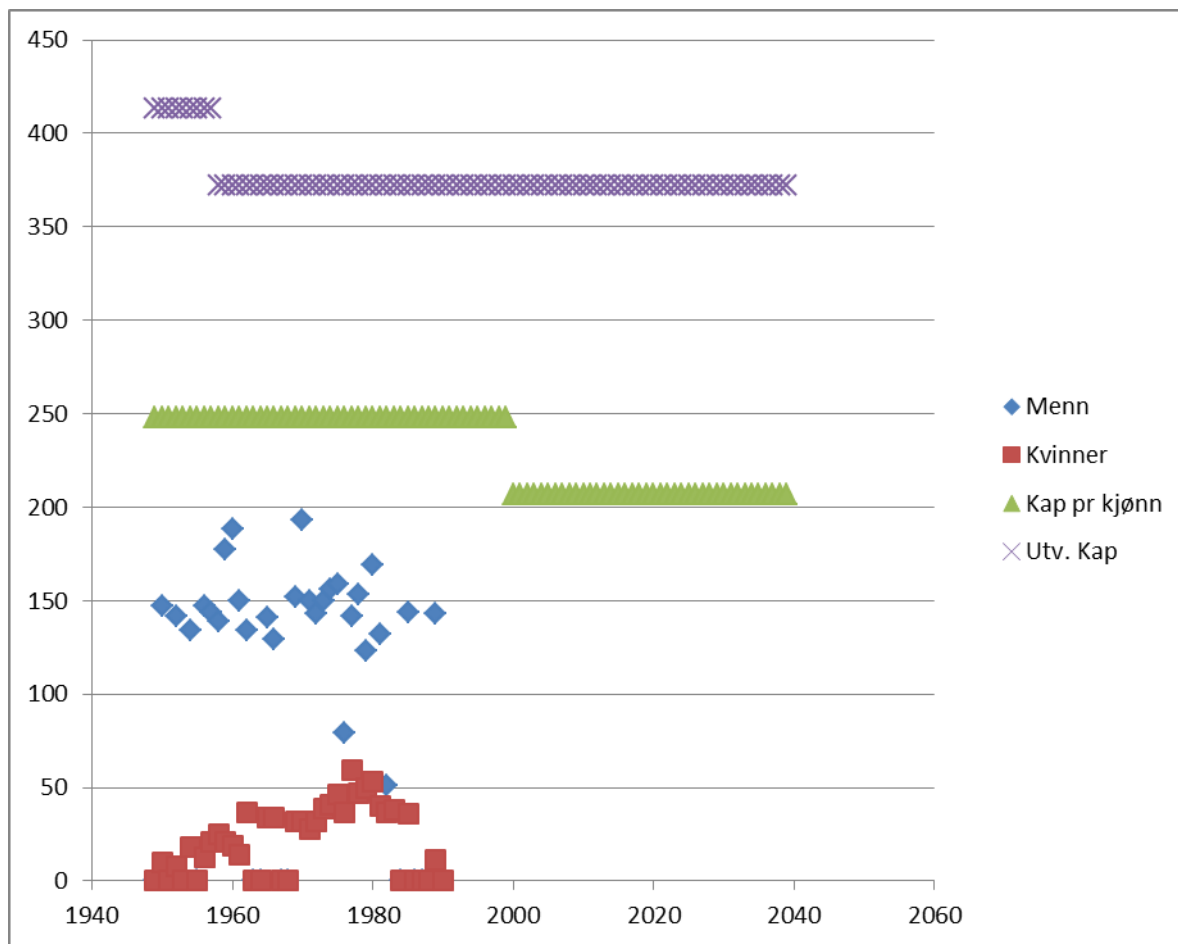


Figur 3 D-nummer pr. år. Menn/ kvinner angir fødselsdato. Kapasitet er gitt pr. fødselsår (under forutsetning av jevn distribusjon over alle fødselsdatoer og begge kjønn.) Nye angir nye tildelinger av D-nummer pr. år, og svarer til samlet tildeling for begge kjønn.

For å vurdere behovet for kapasitet i vid forstand, er det nødvendig til enhver tid å kjenne fødselsårs- og kjønnsfordelingen i eksisterende D-nummerpopulasjon. For eksempel, dersom det på noe tidspunkt (la oss si i år T) inntreffer at antall personer av kjønn S og født i år F som er tildelt D-nummer er lik kapasiteten i vid forstand (for år F), så er kapasiteten sprengt og det kan ikke tildeles D-nummer til en eventuell ny person av kjønn S og med fødselsår F , selv om det benyttes «vedtaksdato» (innenfor samme år.)

La $D_T(F, S)$ være antall personer av kjønn S født år F som er tildelt D-nummer siden ordningen startet (i 1964), observert i (utgangen av) år T . Figur 3 viser $D_{2011}(F, menn)$ (blå kurve) og $D_{2011}(F, kvinner)$ (rød kurve). For eksempel har den blå kurven et maksimum på 25664 menn, født i år 1966, mens den røde kurven har et maksimum på 9276 kvinner født i 1989.

Figur 4 viser **status** for kapasitet i snever forstand (det vil si pr. kjønn og fødselsdag) ved utgangen av 2011. Her er det vist tall for *kritiske datoer*, det vil si de datoene som har flest individer i D-nummerpopulasjonen. For de «kritiske» datoene ligger antallet på ca 1% av størrelsen på populasjonen for tilsvarende fødselsår. Det vil si at de kritiske datoene er overrepresentert med en faktor ca. 3 – 4 i forhold til en uniform distribusjon over årets 365 dager (som skulle tilsi ca. 0,027 % for hver dag.)



Figur 3 Kritiske fødselsdatoer for noen fødselsår. I de fleste tilfeller er det snakk om 1.januar. Spesielt for aldersgruppen født etter 1965 kan vi vente at disse tallene øker, men problemet kan omgås med administrativ tildeling av fiktiv «vedtaksfødselsdato».

Som Figur 4 viser, er kapasiteten i snever forstand fortsatt ikke sprengt. Men i lys av at populasjonen for de kritiske årskullene (og senere årskull) vil øke i årene som kommer, er det sterk grunn til å tro at det vil inntreffe et behov for å tildele «vedtaksdato» etter hvert.

2.2 Beregning av fremtidig behov for D-numre i forhold til kapasitet i vid forstand

La $\tau(T)$ være antall personer totalt som får eller har fått tildelt D-nummer i år T , la $\tau(T, S)$ være antall personer av kjønn S som får eller har fått tildelt D-nummer i år T eller tidligere, og la $\tau(T, F, S)$ være antall personer av kjønn S født år F som får eller har fått tildelt D-nummer i år T eller tidligere.

$D_T(F, S)$, som vi innførte i forrige seksjon, er en ikke-avtagende funksjon av T , på grunn av relasjonen

$$D_T(F, S) = \sum_{t=1964}^T \tau(t, F, S) \quad (1)$$

Dersom verdien av $D_T(F, S)$ noen gang, la oss si i år T , for et fødselsår F og for et kjønn S overstiger kapasiteten (i vid forstand) målt i D-nummer pr. kjønn pr. år, kan vi anse systemet som sprengt. Vi vil betrakte funksjonen

$$D(F, S) = \lim_{T \rightarrow \infty} D_T(F, S) \quad (2)$$

som den *kritiske* i forhold til kapasiteten i vid forstand.

Både aldersfordeling og kjønnsfordeling ved tildelingstidspunkt for D-nummer er svært skjeve. La $f_T(\mathcal{A}, S)$ være den relative frekvensen for alder $\mathcal{A}$ og kjønn S i tildelingsåret. Det vil si,

$$f_T(\mathcal{A}, S) = \frac{\text{Antall D-nummer tildelt i år } T \text{ til personer av kjønn } S \text{ født år } T - \mathcal{A}}{\tau(T)} \quad (3)$$

I prinsippet vil fordelingen $f_T(\mathcal{A}, S)$ variere tilfeldig fra år til år, og underliggende trender kan forandre fordelingen systematisk over tid. For enkelhets skyld kan vi ikke gå inn i detaljene i slike grunnleggende forandringer. Statistikk [4] viser imidlertid en forbløffende likhet i fordelingene for de tre siste årene med kjente data, det vil si at $f_{2009}(\mathcal{A}, S) \approx f_{2010}(\mathcal{A}, S) \approx f_{2011}(\mathcal{A}, S)$, og vi vil anta i videre analyse at $f_T(\mathcal{A}, S)$ er en konstant fordeling $f(\mathcal{A}, S)$ som er uavhengig av T og gitt ved gjennomsnittet av $f_{2009}(\mathcal{A}, S)$, $f_{2010}(\mathcal{A}, S)$ og $f_{2011}(\mathcal{A}, S)$. Se Appendiks B.

Betydningen av fordelingen $f_T(\mathcal{A}, S)$ er at

$$\tau(T, F, S) = \tau(T) \cdot f_T(T - F, S). \quad (4)$$

Ettersom scenariene gir anslag for $\tau(T)$, kan vi på denne måten beregne et anslag for $\tau(T, F, S)$ og dermed, ved iterasjon, verdien av $D_T(F, S)$ for år T . Forutsetningene er altså at vi kan

- A. anslå fremtidige verdier for $\tau(T)$, og
- B. anta at alders- og kjønnsfordelingen $f_T(\mathcal{A}, S)$ holder seg konstant.

Forutsetning A er mest kritisk, og bestemmes av **scenariene**.

Dersom årlig tilvekst i D-nummer, $\tau(T)$, vokser ut over kapasiteten i vid forstand, vil et (informasjonsbærende) personidentifikatorsystem kollapse. Dersom $\tau(T)$ stabiliserer seg («flater ut»), kan vi observere at funksjonen $D_T(F, S)$ vil nærme seg sin endelige verdi for $T - S > 70$. Det kommer av at, som vist i Appendiks B, $f(\mathcal{A}, S)$ er liten i tallverdi for $\mathcal{A} > 70$. I tekst betyr dette at det blir delt ut relativt svært få D-nummer til personer over 70 år. Dermed har vi at grenseverdien

$$D(F, S) \approx D_{F+70}(F, S) \quad (5)$$

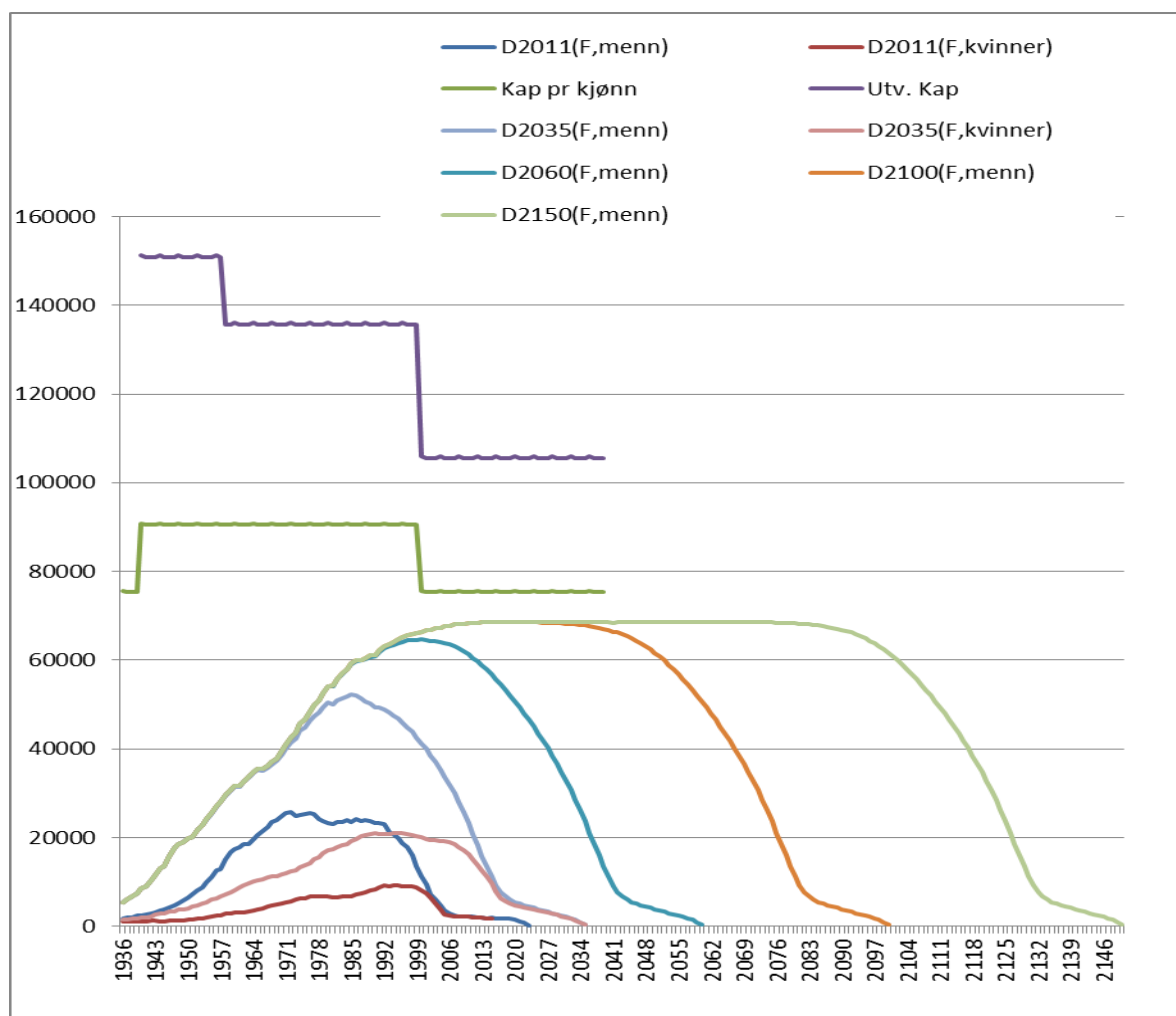
Praktisk metode:

I Appendiks A diskuterer vi forskjellige scenarier (I, II og III, fra forespørselen [3]) for forutsetning A ovenfor. For hvert scenario beregner vi et anslag for den asymptotiske verdien $D(F, S)$ gitt ved (5). Dette uttrykker behovet for en gitt kapasitet i *vid* forstand.

2.3 Kapasitet i vid forstand: Resultater

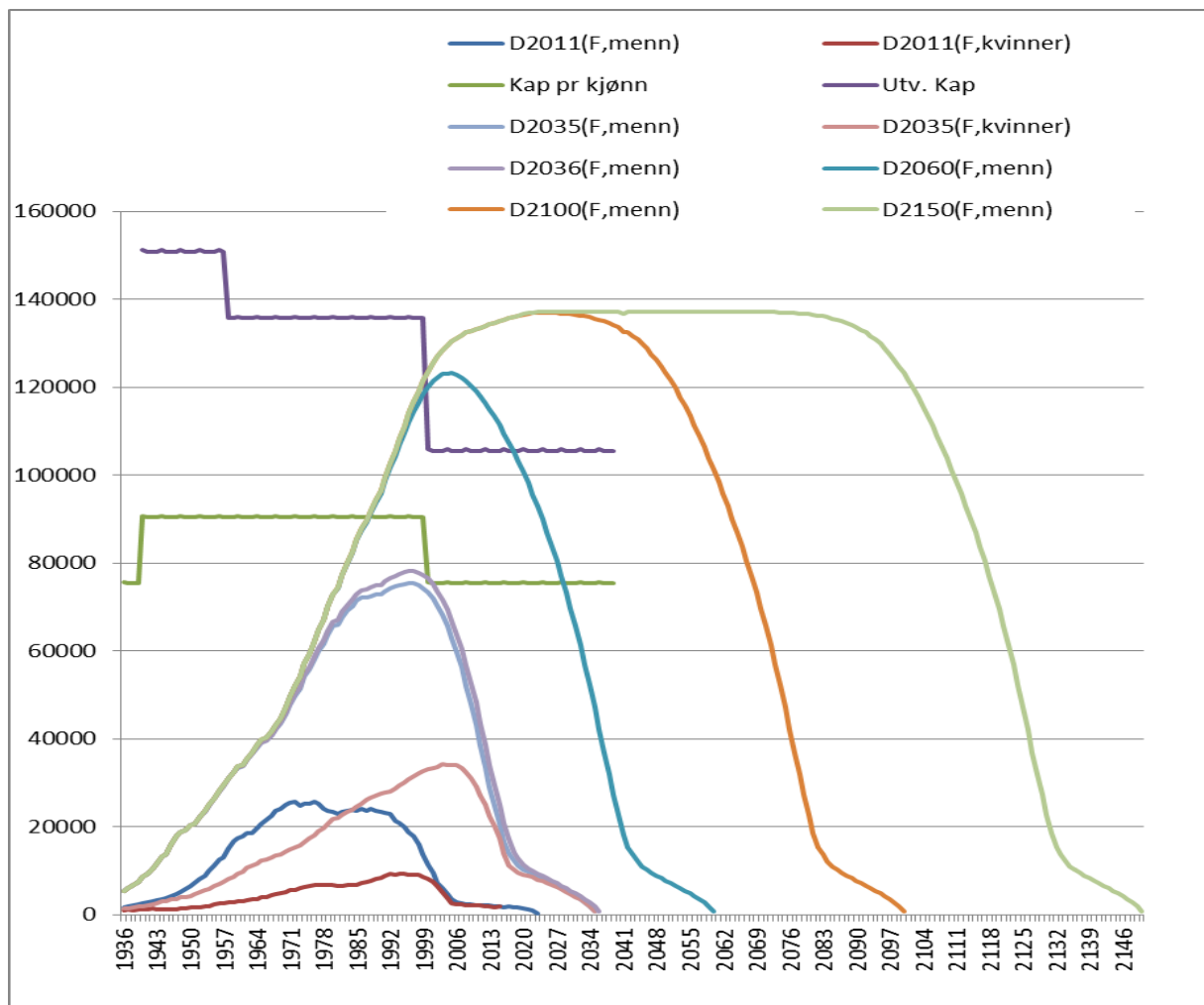
Konsekvenser av Scenariene I, II og III er gitt i Figur 5-7. Figurene viser $D_T(F,S)$ for et utvalg av observasjonsår T og for et utvalg av kjønn S .

Advarsel: Vi presiserer at figurene er gitt i forhold til Scenariene I, II og III og antagelsen om at fordelingsfunksjonen er fast og lik gjennomsnitt av den tilsvarende fordelingen som observert i 2009-2011. Vi assosierer ingen sannsynlighet med disse antagelsene.



Figur 4 Vekst ved **scenario I**. Kapasitet i vid forstand. Fødselsår F langs horisontalaksen.

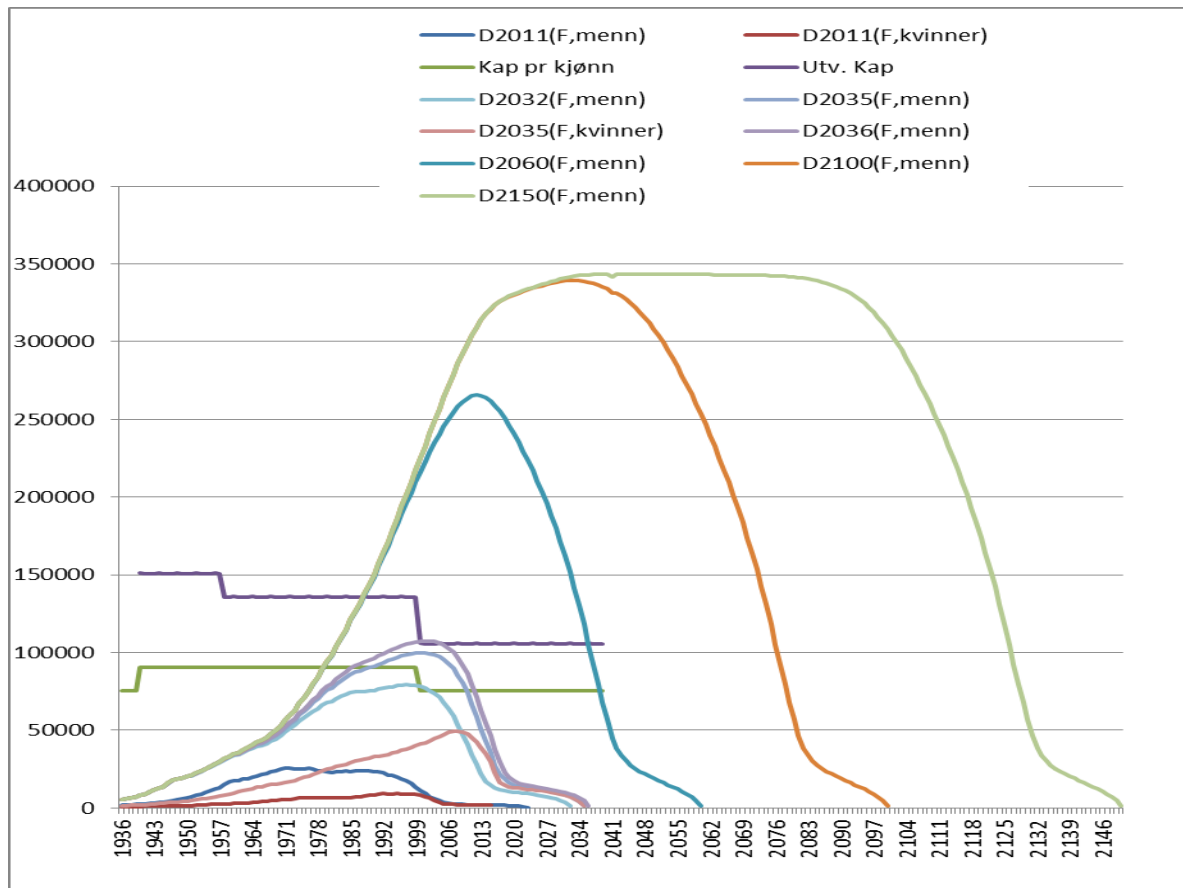
For **scenario I** vil behovet for D-nummer ikke overstige kapasiteten i vid forstand før 2040 (da vil man være nødt til å etablere en ny personidentifikator uansett.) Dette er ikke overraskende; scenariet forutsetter en tilvekst på ca. 68600 menn og 31400 kvinner pr. år, som er det nivået fødselsårskullene stabiliserer seg på.



Figur 5 Vekst ved **Scenario II**. Kapasitet i vid forstand. Fødselsår F langs horisontalaksen.

For scenario II vil behovet for D-nummer overstige kapasiteten i vid forstand.

Mer presist: Til og med 2035 vil behovet, ved **Scenario II**, være oppfylt av kapasiteten i vid forstand (det vil si ordinær kapasitet, ikke utvidet kapasitet) for alle fødselsår og begge kjønn – såvidt det er. I 2036 vil fødselsårskullet 2000 såvidt overstige kapasiteten i vid forstand (slik den er definert i dag), men ikke utvidet kapasitet. På lengre sikt ser vi at også utvidet kapasitet, i vid forstand, vil bli overskredet.



Figur 6 Vekst ved **scenario III**. Kapasitet i vid forstand. Som vi ser vil behovet pr. år stabilisere seg på ca. 70% (det vil si herreandelen) av det totale behovet for D-numre, som man kan forutse på forhånd. Fødselsår F langs horisontalaksen.

Til og med 2031 vil behovet, ved **Scenario III**, være oppfylt av kapasiteten i vid forstand (det vil si ordinær kapasitet, ikke utvidet kapasitet) for alle fødselsår og begge kjønn – såvidt det er. I 2032 vil fødselsårskullet 2000 såvidt overstige kapasiteten i vid forstand.

Til og med 2035 vil behovet, ved Scenario III, være oppfylt av **utvidet** kapasitet i vid forstand for alle fødselsår og begge kjønn – såvidt det er. I 2036 vil fødselsårskullene 2000-2004 i størrelse såvidt overstige kapasiteten i vid forstand.

2.4 Kapasitet i snever forstand

For å beregne kapasiteten i *snever* forstand (det vil si per dag) kan vi studere forholdet mellom antall individer født på kritiske dager for hvert år (Figur 4), i forhold til verdiene $D_{2011}(F, menn)$ for de samme fødselsårene F . Ved å anta at dette forholdet holder seg konstant, kan vi få et (grovt!) anslag for fremtidige behov for D-nummer pr. dag. En slik analyse vil typisk gi som konklusjon at dagens D-nummerordning *ikke* vil kunne tilby kapasitet i snever forstand som kan oppfylle fremtidige behov, i henhold til scenariene I, II og III.

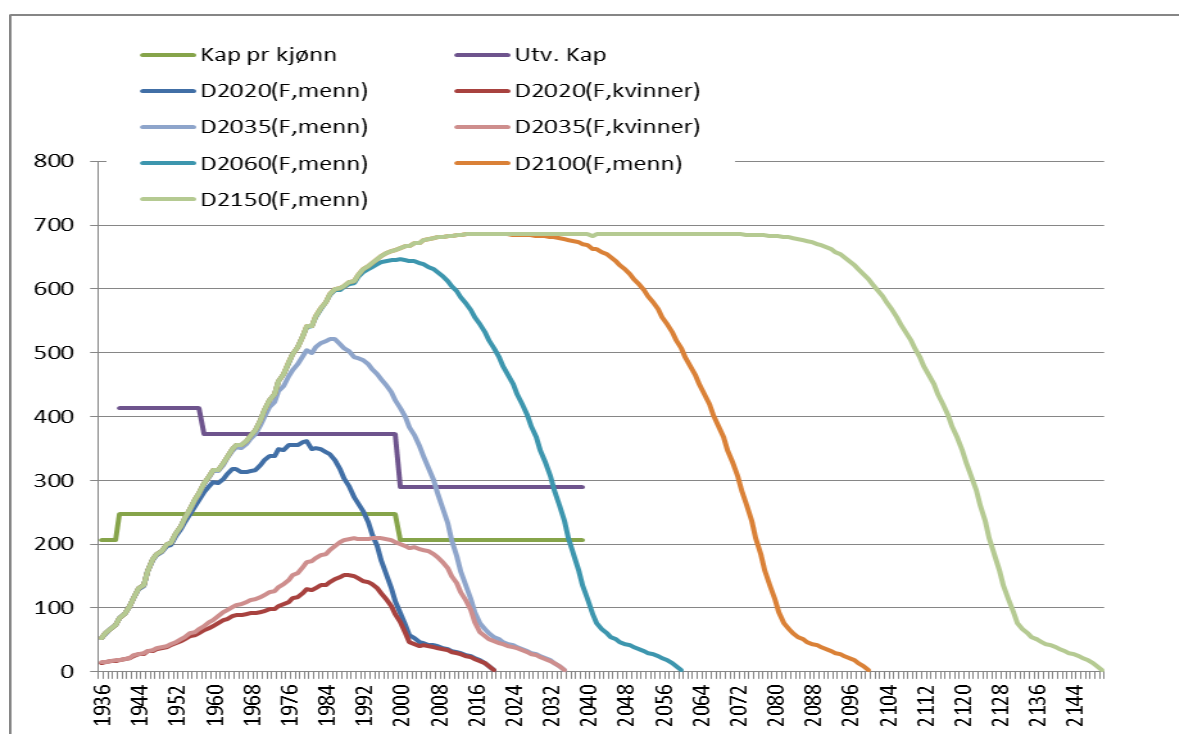
På grunn av tilfeldig (og også systematisk) variasjon i fødselsdato, er det vanskeligere å oppfylle behovet for kapasitet i snever forstand (enn i vid forstand). Resultatene i seksjon 2.3 indikerer derfor allerede at D-nummerordningen ikke vil møte behovet i snever forstand.

Mer konkret viser dataene at i de mest ekstreme fødselsårene faller i overkant av 1% av fødselsdatoene på kritiske dager. For menn er verste dag så langt 1.januar 1970, med 1.07% av alle D-numre tildelt menn *født* i 1970. Dette tilsvarer 1.07% prosent av alle D-numre tildelt menn i 1970. Legg merke til at 1.07% tilsvarer omtrent 4 ganger $1/365 \approx 0.0027$. Det betyr at skjevfordelingen av fødselsdato alene gjør at behovet for kapasitet i snever forstand blir «fire ganger så høyt» som kapasitetsbehovet i vid forstand.

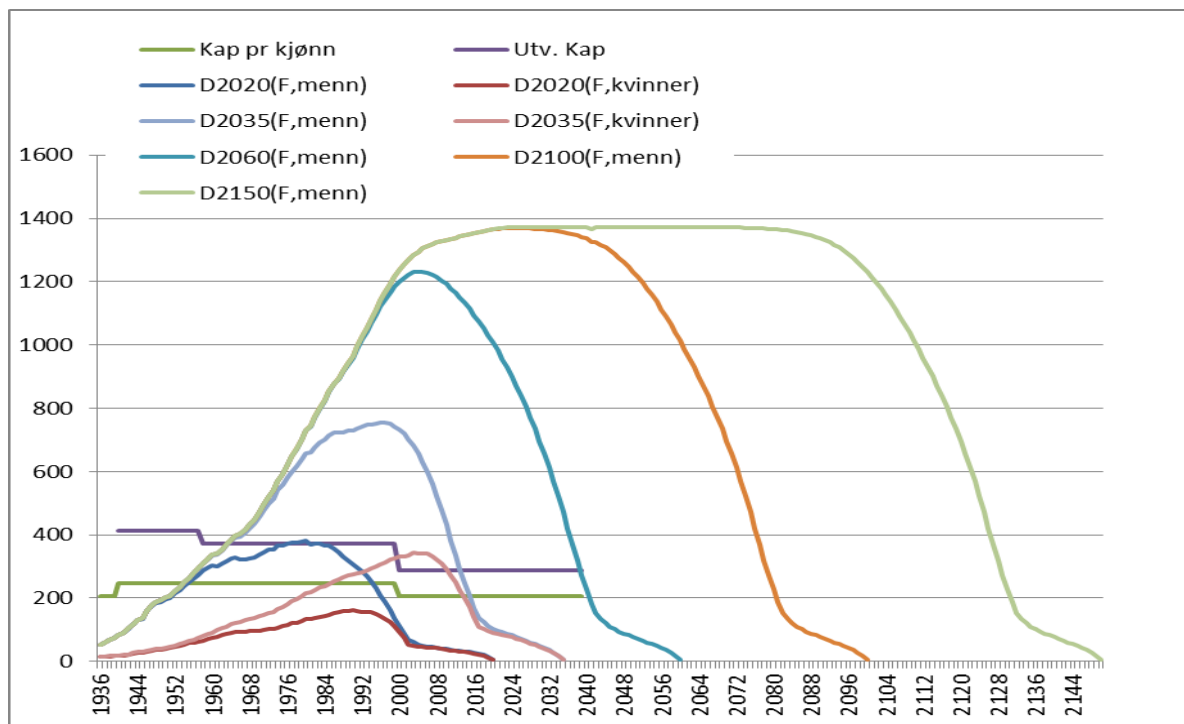
For kvinner er fødselsdatofordelingen typisk enda skjevare (opp til over 2% av kvinnepopulasjonen født i 1977 har offisiell fødselsdag 1.januar), men her er populasjonen mindre og konsekvensen blir også mindre.

Figur 8-10 viser behov i forhold til kapasitet i snever forstand, tilsvarende Figur 5-7. Disse figurene viser at selv ved **Scenario I** og selv ved utvidet kapasitet i dagens D-nummerordning som skissert i Kapittel 1), vil kapasiteten i snever forstand ikke oppfylle behovet, og det vil bli nødvendig å ta i bruk vedtaksdato (i ca. 1% av nye tilfeller) ganske snart.

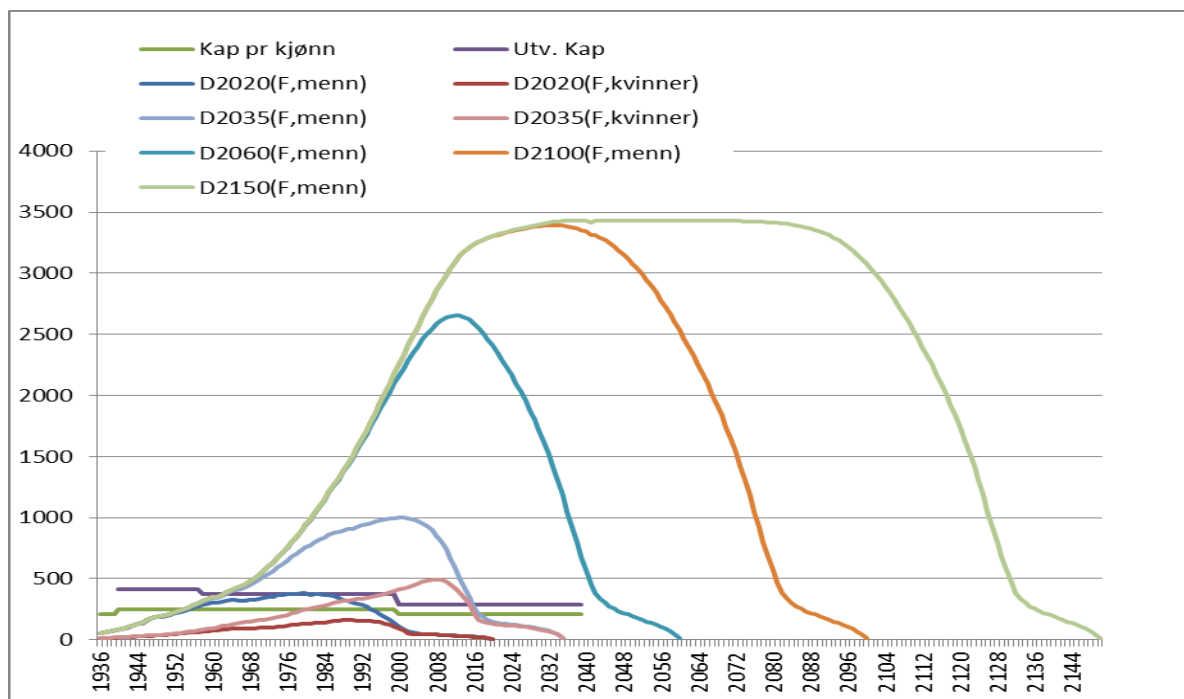
Forbehold: I Figur 8-10 har vi antatt at 1% av årskullet i D-nummerpopulasjonen samler seg på en felles «kritisk» fødselsdato. Selv om dette samsvarer med data i eksisterende D-nummerpopulasjon, er det ikke gitt at dette forholdet vil vedvare i fremtiden, og vi assosierer ingen sannsynlighet med denne antagelsen. Forøvrig gjelder de samme forbehold som under diskusjon av kapasitet i vid forstand som vist i Figur 5-7.



Figur 7 Vekst ved **scenario I**. Kapasitet i snever forstand: Kapasitet og behov pr. kjønn og dag. Fødselsår F langs horisontalaksen. Under forutsetning av at 1 % av D-nummertildelinger til personer født i år F samles på kritiske fødselsdager, vil behovet for D-nummer for menn overstige kapasiteten i snever forstand. Altså vil det bli nødvendig med «vedtaksdato»-tildelinger. **Modellen** tilsier at for noen årskull vil det skje allerede fra 2012, men i realiteten er vi bare «nesten der» - dvs. for noen årskull er behovet oppunder 200 menn pr. dag. Dette gjelder årskull som ikke er de største i populasjonen. Merk at de kritiske årskullene vil vokse de nærmeste årene.



Figur 8 Vekst ved **scenario II**. Kapasitet i snever forstand: Kapasitet og behov pr. kjønn og dag. Fødselsår F langs horisontalaksen



Figur 9 Vekst ved **scenario III**. Kapasitet i snever forstand: Kapasitet og behov pr. kjønn og dag. Fødselsår F langs horisontalaksen.

3) Robusthet for fødselsnummer ved økt populasjon/personkrets

Dette kapittelet er et svar på følgende oppgave fra [3]:

Oppgave 3

Oppgaven er å beregne alternativenes kapasitet/robusthet ved 3 ulike alternativer for utvidelse av personidentifikatorens populasjon. De tre alternativene er:

- Personkretsen til D-nummeret inngår i en framtidig personidentifikator.
- Personkretsen til D-nummeret og DUF-nummeret inngår i en framtidig personidentifikator.
- Personkretsen til D-nummeret, DUF-nummeret og helse-nummeret inngår i en framtidig personidentifikator.
- Personkretsen til de som er sikkert identifisert gjennom ID-kontroll inngår i en framtidig personidentifikator. SKD spiller inn hvor stor denne gruppen er.

Kommentar: med hensyn til de siste tre alternativene: Personkretsen til D-nummeret og DUF-nummeret inngår i en framtidig personidentifikator, Personkretsen til D-nummeret, DUF-nummeret og helse-nummeret inngår i en framtidig personidentifikator., Personkretsen til de som er sikkert identifisert gjennom ID-kontroll inngår i en framtidig personidentifikator. SKD spiller inn hvor stor denne gruppen er...så mangler vi pålitelige prognoser for vekst i disse. Vi har derfor ikke vurdert kapasitetsbehovet for disse alternativene. Se diskusjon i Appendiks A.

Nedenfor diskuterer vi det første alternativet: Personkretsen til D-nummeret inngår i en framtidig personidentifikator. Hovedkonklusjonen vår er at alle fire alternativ vil ha tilstrekkelig kapasitet, i enhver forstand og i forhold til alle vurderte vekstscenarier, fram til 2100. På grunn av manglende prognoser for befolkningsvekst utover 2100 har vi ikke kunnet vurdere kapasitet i forhold til behov fram til 2150. En rimelig ekstrapolasjon av vekstscenariene fram til 2150 tilsier de informasjonsbærende Alternativ 1 og 2 vil tilfredsstille kapasitetsbehovet i vid forstand, men ikke i snever forstand, frem til 2150. Under samme forutsetninger kan man anta at de informasjonsløse Alternativ 3 og 4 vil tilfredsstille behovene fram til 2150.

De fire aktuelle alternativene for ny personidentifikator som er gjenstand for analyse er:

	Informasjonsbærende	Informasjonsløs
Numeriske tegn	0. Dagens fødselsnummer: DDMMÅÅIIKK. 1. Dagens løsning med endret kontrollberegning: DDMMÅÅIIKK.	3. Kundenummer: IIIIIIIKK. Har to varianter: 3a) Kun nye borgere 3b) Hele befolkningen skifter.
Alfanumeriske tegn	2. Informasjonsbærende og alfanumerisk. DDMMÅÅAAKK.	4. Informasjonsløs og alfanumerisk (8): AAAAAAKK. Hele befolkningen skifter.

K – kontrollsiffer, I – Individsiffer, A – Alfnumeriske tegn

Med hensyn til kapasitet for de ulike alternativene, henviser vi til Kapittel 1 (Innledningen), Kapittel 5, og til [1] og [2]. Robusthet for de forskjellige alternativene avhenger av de forskjellige scenariene.

Metode

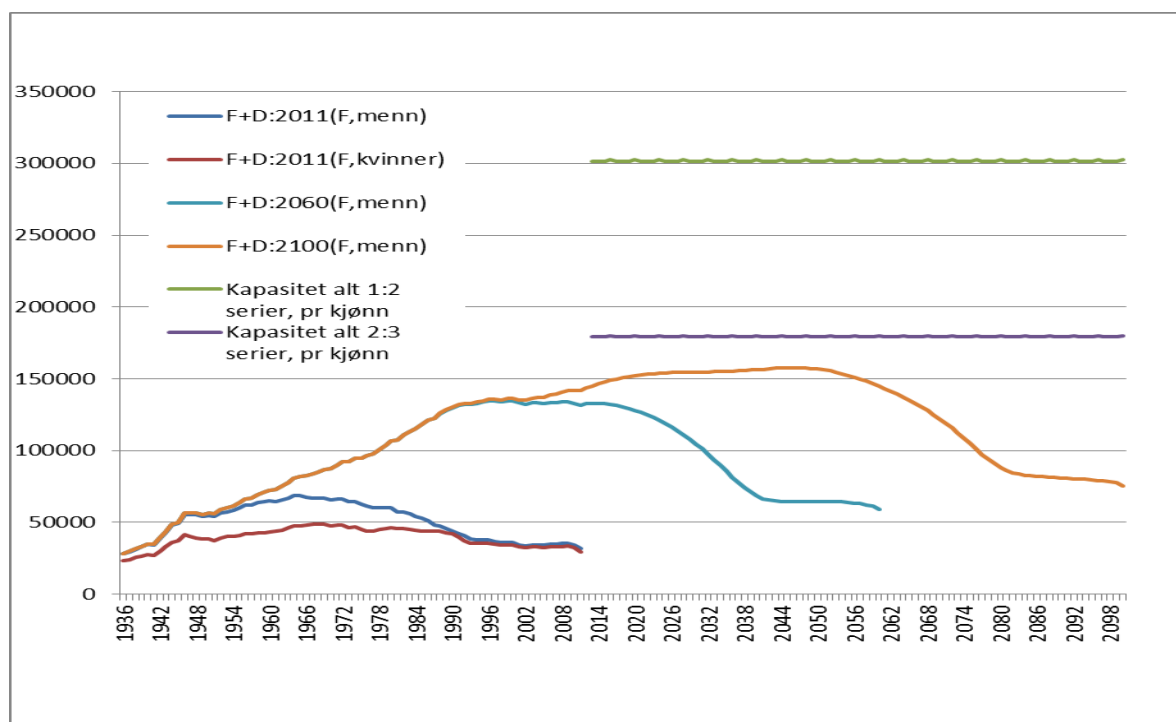
For å estimere størrelse og aldersfordelinger på populasjoner, bruker vi estimer for befolkningsvekst i henhold til scenarier og prognoser (Appendiks A) og estimer for alders- og kjønnsfordeling på tildelingstidspunkt (Appendiks B).

Kapasitet for Alternativ 1 og Alternativ 2

Alternativ 1 og Alternativ 2 [1] har kapasitet på henholdsvis 8264 og 7535 pr. dag, eller i henholdsvis 3.016.360 og 2.750.275 pr. år. Denne kapasiteten kan porsjoneres ut i henholdsvis tiendedeler og 1/23-deler, for koding av århundre og kjønn. Tar vi kjønn i betraktning og går ut fra en *jevn fordeling av kapasiteten på de to kjønn*, har Alternativ 1 og Alternativ 2 kapasitet i snever forstand henholdsvis 4132 og 3767 pr. dag pr. kjønn, eller i vid forstand henholdsvis 1.5 millioner og 1.37 millioner pr. år pr. kjønn.

Kapasitet i vid forstand : D+F-nummer for Alternativ 1 og Alternativ 2

Figur 11 viser aldersfordeling i den kombinerte D-nummer+Fødselsnummerpopulasjonen pr årskull i henhold til vekstscenariet «høy nasjonal vekst» og D-nummer-vekstscenario I.

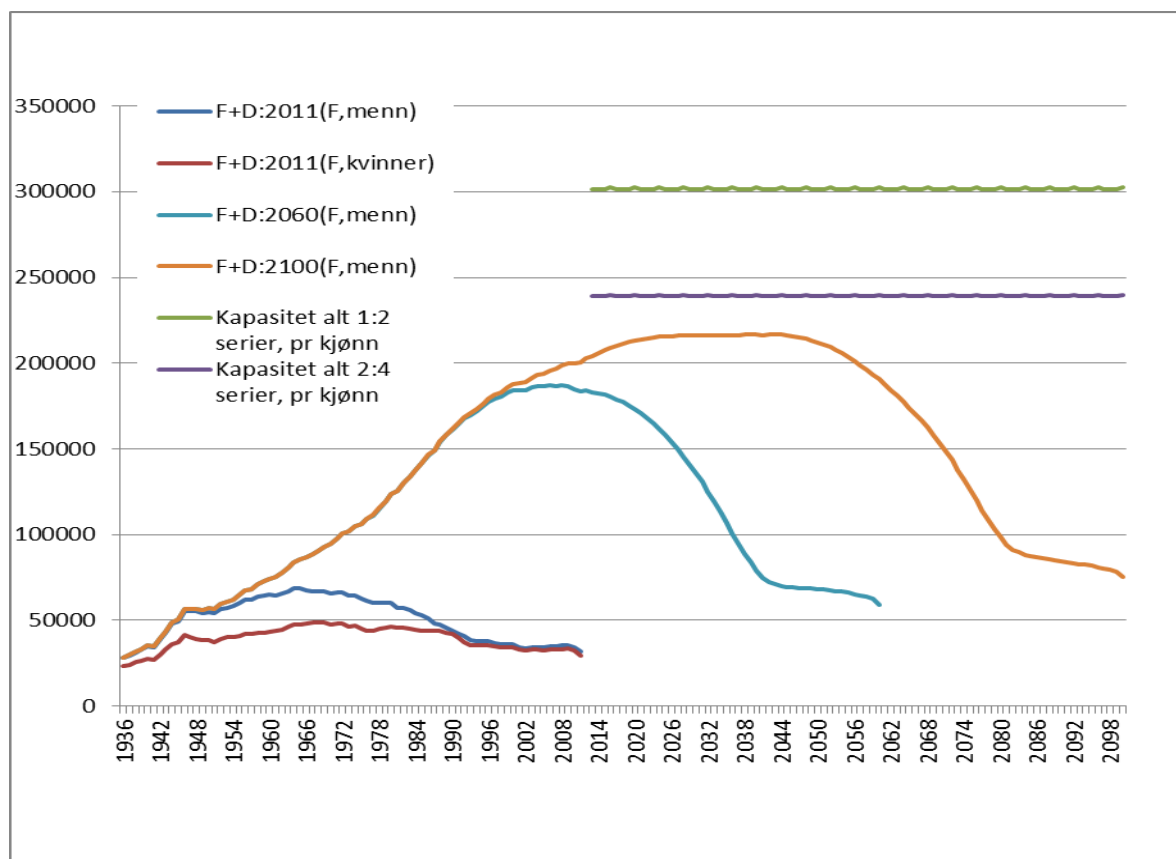


Figur 11: Alders- og kjønnsfordeling i D-nummer+Fødselsnummerpopulasjon pr årskull i henhold til vekstscenariet høy nasjonal vekst, samt D-nummer-vekstscenario I. Antagelser som i Appendiks A og B. Fødselsår F langs horisontalaksen.

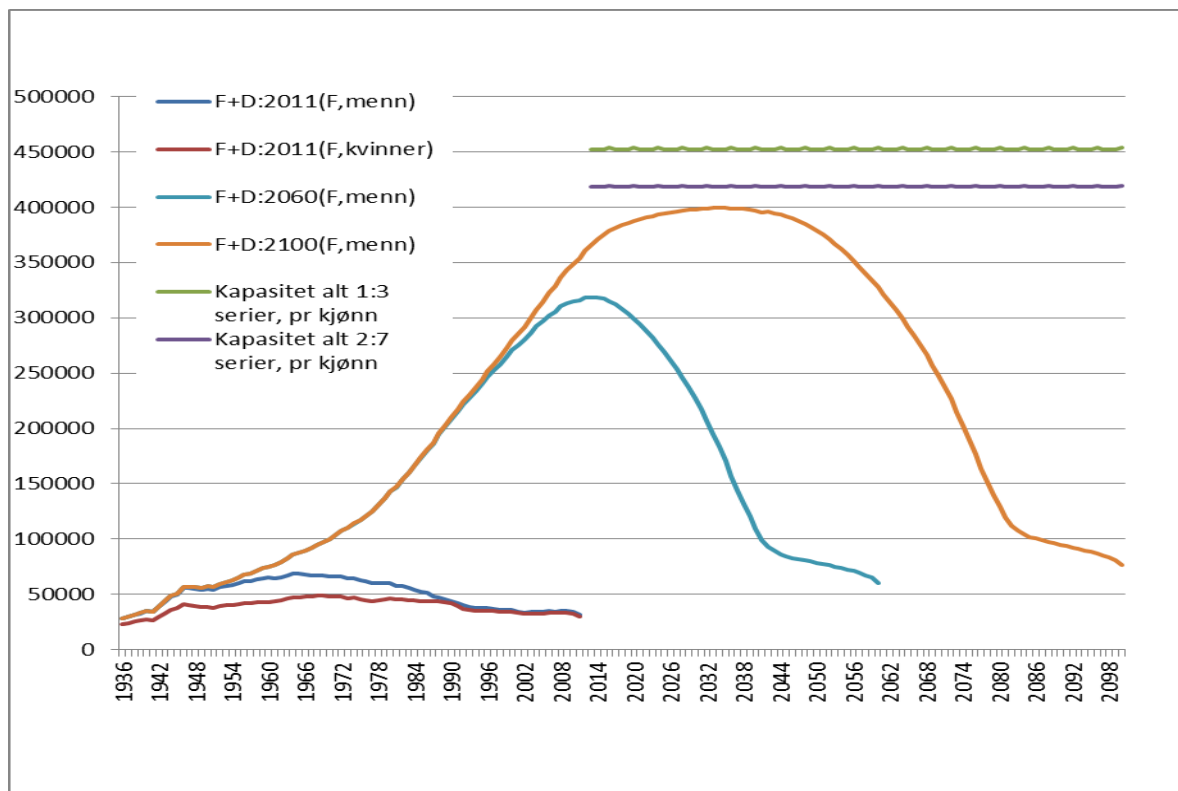
Legg merke til at i dette scenariet (og dersom vi forutsetter en jevn fordeling av kapasiteten delt på menn og kvinner) er det nødvendig å bruke to av 10 tilgjengelige nummerserier for Alternativ 1 i perioden frem til 2100 (Dersom nummerserier koder for fødselsårhundre, betyr det i praksis at disse to nummerseriene vil gjelde fra starttidspunkt for ordningen og frem til starttidspunkt +100 år). For alternativ 2 må det brukes tre av 23 tilgjengelige nummerserier.

Figur 12 viser aldersfordeling i den kombinerte D-nummer+Fødselsnummerpopulasjonen pr årskull i henhold til vekstscenariet «høy nasjonal vekst» og D-nummer-vekstscenario II. I dette scenariet er det fremdeles tilstrekkelig å bruke to nummerserier for Alternativ, men det er nødvendig å bruke fire nummerserier for Alternativ 2.

Figur 13 viser aldersfordeling i den kombinerte D-nummer+Fødselsnummerpopulasjonen pr årskull i henhold til vekstscenariet «høy nasjonal vekst» og D-nummer-vekstscenario III. Dette vekstscenariet krever tre nummerserier av 10 tilgjengelige for Alternativ 1, og 7 av 23 tilgjengelige for Alternativ 2.



Figur 12 : Alders- og kjønnsfordeling i D-nummer+Fødselsnummerpopulasjon pr årskull i henhold til vekstscenariet høy nasjonal vekst, samt D-nummer-vekstscenario II. Antagelser som i Appendiks A og B. Fødselsår F langs horisontalaksen.

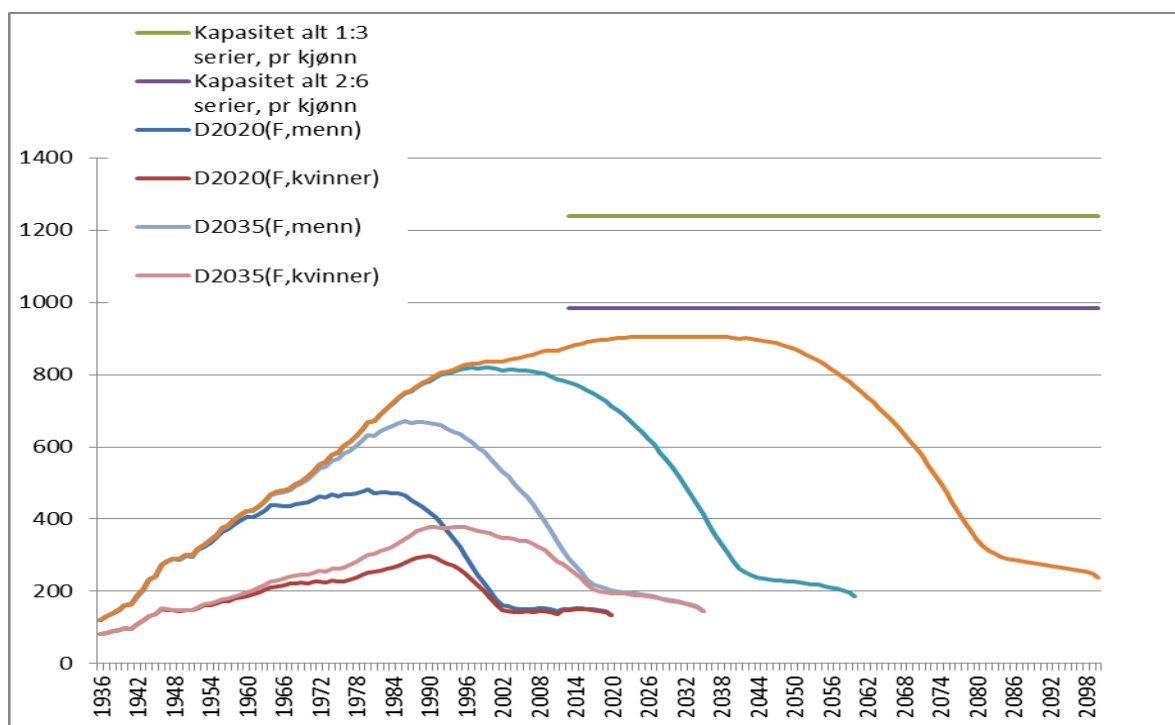


Figur 13 : Alders- og kjønnsfordeling i D-nummer+Fødselsnummerpopulasjon pr årskull i henhold til vekstscenariet høy nasjonal vekst, samt D-nummer-scenario III. Antagelser som i Appendiks A og B. Fødselsår F langs horisontalaksen.

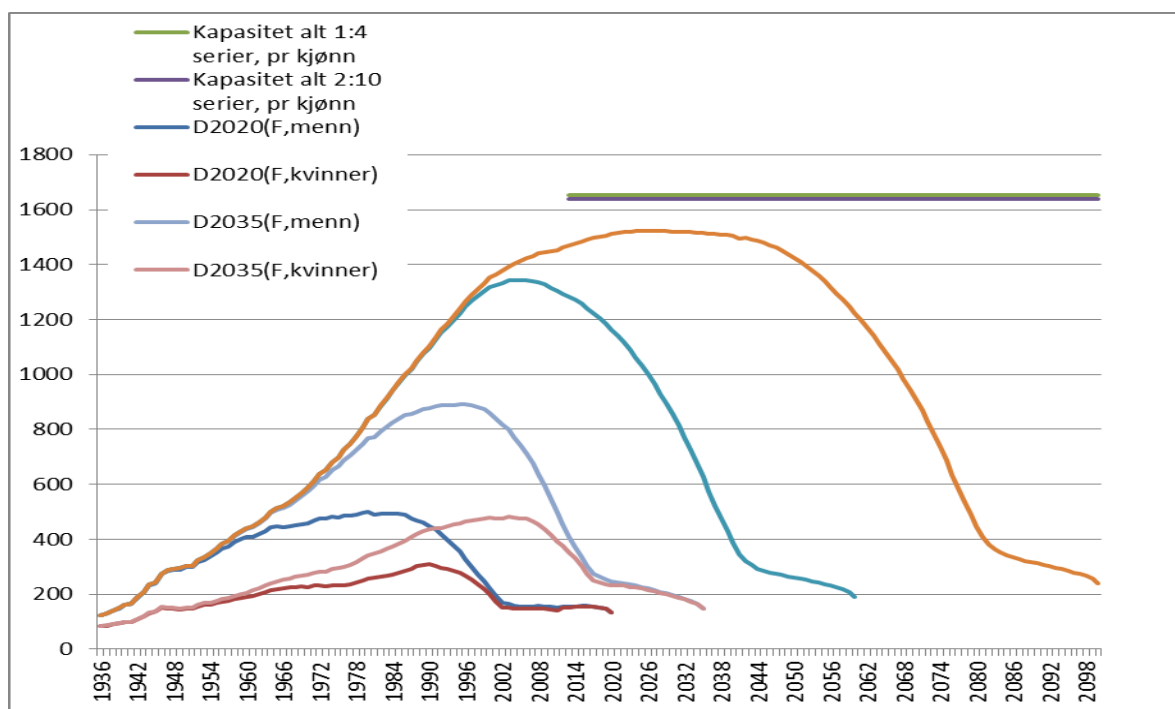
Kapasitet i snever forstand : D+F-nummer for Alternativ 1 og Alternativ 2

I Kapittel 2 har vi redegjort for hvorfor vi i D-nummerpopulasjonen har vi brukt 1% av årskullet som et estimat for antall som samles på «kritiske dager». Fødselsnummerpopulasjonen er mer homogen, og Statistisk Sentralbyrå anslår (se Appendiks 5 i [1]) at de kritiske dagene har 14% høyere antall individ enn gjennomsnittsdagen (det vil si, $1.14 * 1/365$ av årskullet.) Disse to anslagene er ikke nødvendigvis konsistente i en situasjon der en stor del av befolkningsveksten inntreffer på grunn av innvandring.

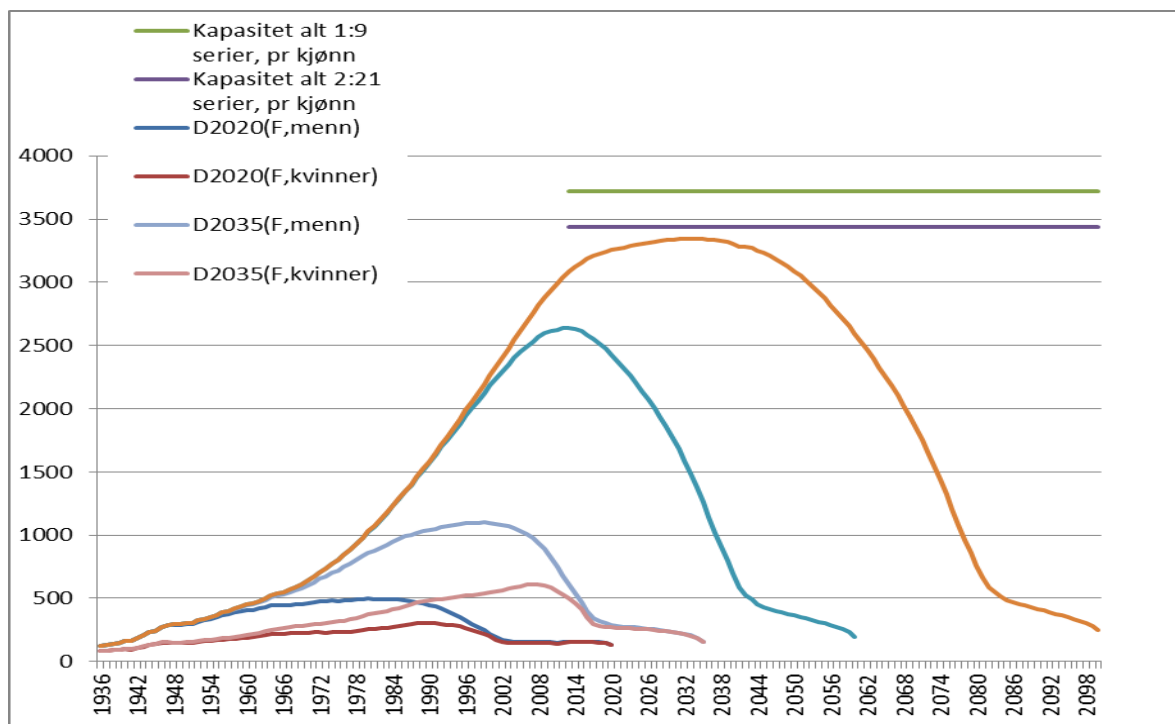
I mangel av pålitelige prognoser har vi valgt en modell der vi betrakter summen av antall individer på kritisk dag i D-nummerpopulasjonen, anslått til $a = 1\%$ av D-nummerårskullet* 90% (fordi 10% av årskullet regnes inn i fødselsnummerpopulasjonen på grunn av overlapp, se Appendiks A), og summen av antall individer på kritisk dag i fødselsnummerpopulasjonen, anslått til $b = 3.12\%$ ($= 1.14/365$) av fødselsnummerårskullet. Vi antar at kritisk dag er den samme (typisk 1.januar) i begge tilfeller, og bruker $a+b$ som anslag for antallet individer født på kritisk dag.



Figur 14 Alders- og kjønnsfordeling i D-nummer+Fødselsnummerpopulasjon pr årskull og kritiske dag i henhold til vekstscenariet høy nasjonal vekst, samt D-nummer-scenarior I. Antagelser som i Appendiks A og B. Fødselsår F langs horisontalaksen, antall pr. kritisk dag langs vertikalaksen.



Figur 10 Alders- og kjønnsfordeling i D-nummer+Fødselsnummerpopulasjon pr årskull og kritiske dag i henhold til vekstscenariet høy nasjonal vekst, samt D-nummer-scenarior II. Antagelser som i Appendiks A og B. Fødselsår F langs horisontalaksen, antall pr. kritisk dag langs vertikalaksen.



Figur 11 Alders-og kjønnsfordeling i D-nummer+Fødselsnummerpopulasjon pr årskull og kritiske dag i henhold til vekstscenariet høy nasjonal vekst, samt D-nummer-scenarior I, II og III. Antagelser som i Appendiks A og B. Fødselsår F langs horisontalaksen, antall pr. kritiske dag langs vertikalaksen.

Som figurene viser vil Alternativ 1 takle de kombinerte scenariene «Høy nasjonal vekst»+ «Scenario I» og «Høy nasjonal vekst»+ «Scenario II». For det kombinerte scenariet «Høy nasjonal vekst»+ «Scenario II» vil det da være nødvendig å bruke opp 4 av 10 tilgjengelige nummerserier for å dekke behovet frem til 2100, (som vil angi århundre fra starttidspunkt og 100 år frem i tid) for Alternativ 1, og 9 av 23 tilgjengelige nummerserier for Alternativ 2.

For det kombinerte scenariet «Høy nasjonal vekst»+ «Scenario III» vil det være nødvendig å bruke opp 9 av 10 tilgjengelige nummerserier for å dekke behovet frem til 2100 for Alternativ 1, og 21 av 23 tilgjengelige nummerserier for Alternativ 2.

For det kombinerte scenariet «Høy nasjonal vekst»+ «Scenario III» vil det bli nødvendig å utvide kapasiteten om omtrent hundre år dersom kravet om perfekt informasjonsinnhold skal opprettholdes, det vil si dersom kapasiteten skal tilfredsstille behovet i snever forstand. Alle alternativer (1, 2, 3 og 4) kan øke kapasiteten relativt sømløst (hvis det blir forberedt for det nå) ved å omgjøre et kontrollsiffer til et datasiffer (som det nå blir gjort i Alternativ 1). Med hensyn til Alternativ 1 vil dette medføre å fjerne all gjenværende feilkontroll. De andre alternativene har mer å gå på med hensyn på feilkontroll ved fremtidige utvidelser, og Alternativ 2 og 4 vil også gi en større gevinst i form av økt kapasitet ved en slik utvidelse.

Informasjonsløse identifikatorer

Totalt antall tildelte identifikatorer frem til og med 2100, under forutsetning av høy nasjonal vekst i befolkningen, vil være henholdsvis 34,3 millioner, 41,6 millioner, og 63 millioner for D-nummerscenariene I, II, og III. Merk at de informasjonsløse identifikatorene har kapasitet på henholdsvis **123** millioner (Kap. 5) og **148** millioner [1]. Det betyr at selv om vi forlenger scenariene med tilnærmet samme vekstrate frem til år 2150, bør både Alternativ 3 og Alternativ 4 dekke behovet.

4) Øke kapasitet for D-nummer

Dette kapitlet er et svar på følgende oppgave fra [3]:

Oppgave

Vil det være mulig å øke kapasiteten til dagens D-nummer ved å gjøre enkle grep? For eksempel øke sifrene for måned ved å legge til 2 eller 6 til det tredje nummeret (jamfør helsennummeret hvor det er lagt til 4). Eller ved å endre algoritmen til moduluskontrollen slik dette er foreslått i alternativ 1 for fødselsnummer.

Hva vil være en anbefalt modell for å øke kapasiteten til D-nummeret?

(Dette er en mer omfattende oppgave, og kan gjøres etter at de andre oppgavene er gjennomført.)

I prinsippet vil dagens D-nummersystem kunne endres på samme måte som fødselsnummeret. En utvidelse ved å legge til 2 eller 6 til det tredje sifferet vil også kunne fungere.

1. Legge til 2,6 eller 8 til tredje siffer: Vil kunne gi opp til en firedobling av kapasiteten. Kostnaden ved dette er at fødselsdato vil se enda rarer ut, og det vil også kreve en omprogrammering av programvare for å tolke og verifisere D-numre.
2. Omgjøring av første kontrollsiffer, som i Alternativ 1, vil gi en ellevedobling (hvis dette er et ord) av kapasiteten. Kostnaden ved dette er en svekking av feilkontroll, og det vil også kreve en omprogrammering av programvare for å tolke og verifisere D-numre.

Vi antar at kostnadene ved endring av programvare vil være de samme for disse to alternativene.

Med tanke på forskjellen i kapasitetsøkning, vil vi anbefale omgjøring av første kontrollsiffer.

5) Kapasitet for fødselsnummer alternativ 3

Dette kapittelet er et svar på følgende oppgave fra [3]:

Oppgave

Hvor stor er kapasiteten for alternativ 3 når følgende nummerserier tas bort:

- *Nummerserier brukt som fødselsnummer etter dagens system.*
- *Nummer som kan forveksles med fødselsdato (første siffer 01 – 31).*
- *Nummerserier brukt som D-nummer (første siffer fra 41 – 71).*
- *Nummerserier brukt som Helsenummeret (serier som starter med 8 og 9: 8- og 90000000000 serien. Dette er nummerserier som brukes i framtiden. Fram til i dag har de brukt helsenummer hvor det legges til 4 i tredje siffer – som også bør unngås.*
- *Bankenes Standardiseringskontor (BSK) har oppgitt at følgende nummerserier er i bruk av bankene som konstruerte kundennummer. Konstruerte kundennummer har avtakende bruk slik at det er usikkert hvorvidt dette er et problem på sikt. Foreløpig ønsker banknæringen at man unngår følgende nummerserier:*
 - *00210000000 – 00219999999*
 - *00220000000 – 00229999999*
 - *33000000000 – 33499999999*
 - *32000000000 – 32999999999*
 - *34000000000 – 34499999999*
 - *40000000000 – 40999999999.*

For denne oppgaven er det interessant å synliggjøre frafall av kapasitet til de ulike punktene, dvs. som følge av eksisterende fødselsnummer, D-nummer, Helse-nummer etc.

Fra	til	Utgår pga	Utgår, antall	Gjenstår, antall	Totalt
000.000.000	002.099.999			1.735.536	
002.100.000	002.299.999	Bank	165.288		
002.300.000	009.999.999			6.363.636	
010.000.000	319.999.999	Dato eller Datoaktig	256.198.346		
320.000.000	334.999.999	Bank	12.396.693		
335.000.000	339.999.999			4.132.231	
340.000.000	344.999.999	Bank	4.132.231		
345.000.000	399.999.999			45.454.545	
400.000.000	409.999.999	Bank	8.264.462		
410.000.000	719.999.999	D-nummer eller D-nummeraktig	256.198.346		
720.000.000	799.999.999			66.115.702	
800.000.000	999.999.999	Helsenummer	165.289.255		
			702.644.623	123.801.649	826.446.272

Tabellen angir antall tilgjengelige identifikatorer i de forskjellige nummerseriene. Felt merket **grønt** svarer til nummerserier som inngår som lovlige, felt merket med **rosa** svarer til felt som utgår av forskjellige grupper.

Kommentar: En informasjonsløs identifikator med 11 sifre er vanskeligere å huske, både for individet selv (langtidshukommelse) og for saksbehandlere og andre (korttidshukommelse). Dette øker sannsynlighet for feil. En løsning kan være å utnytte den relativt gode kapasiteten på 123 millioner (pr i dag er det ca 9.4 millioner tildelte fødselsnumre og D-numre) til **midlertidig** å bruke det 9. sifferet til et ekstra kontrollsiffer, i den grad kapasiteten tillater det. Dette vil øke robustheten og minske sannsynligheten for uoppdagete feil.

6) Andre kommentarer

- Vi registrerer at man ikke anser bankkontonummer som en bekymring med tanke på sammenblanding med personidentifikator. I den sammenheng bemerker vi at 11-sifrete bankkontonumre så vidt vi forstår bruker samme kontrollsiffer-vekter for det siste symbolet (k_2) som dagens fødselsnummer, og dermed også som Alternativ 1 og Alternativ 3. Dette medfører at det vil forekomme numre som er både gyldige identifikatorer og gyldige bankkontonumre.
- Til 1.5 i forespørselen [3]: «Kollisjon» med 12 sifre er mulig dersom et symbol ved en feil legges til (for eksempel ved feilaktig gjentakelse av et symbol) i en lengde-11-identifikator, eller dersom et symbol utelates fra et DUF-nummer, - en såkalt «Insertion/deletion»-kanal. I praksis kan man derfor ikke se helt bort fra muligheten for at DUF-numre kan blandes sammen med nye personidentifikator-verdier.

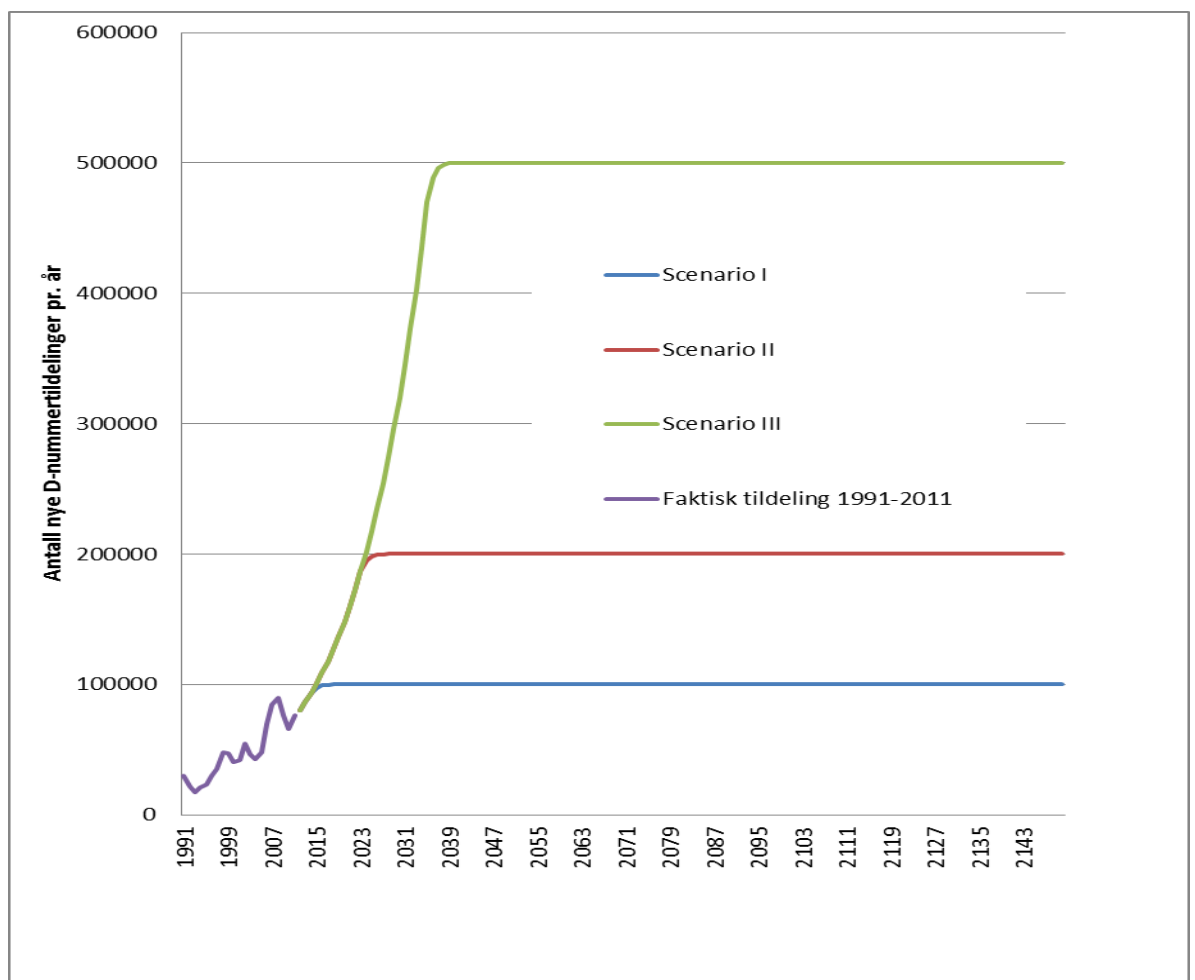
Appendiks A: Vekst i populasjoner: Scenarier

D-nummer-populasjon

Vi har gått ut fra følgende oppgitte scenarier. Vi vil ikke vurdere sannsynlighet for disse scenariene; det avhenger av nasjonal – og globalpolitiske forhold, økonomiske forhold, administrative rutiner, kriger, epidemier og naturkatastrofer, blant annet.

- **Scenario I:** Årlig prosentvis vekst i tilvekst inntil man når 100 000 nye D-nummer pr år.
- **Scenario II:** Årlig prosentvis vekst i tilvekst inntil man når 200 000 nye D-nummer pr år.
- **Scenario III:** Årlig prosentvis vekst i tilvekst inntil man når 500 000 nye D-nummer pr år.

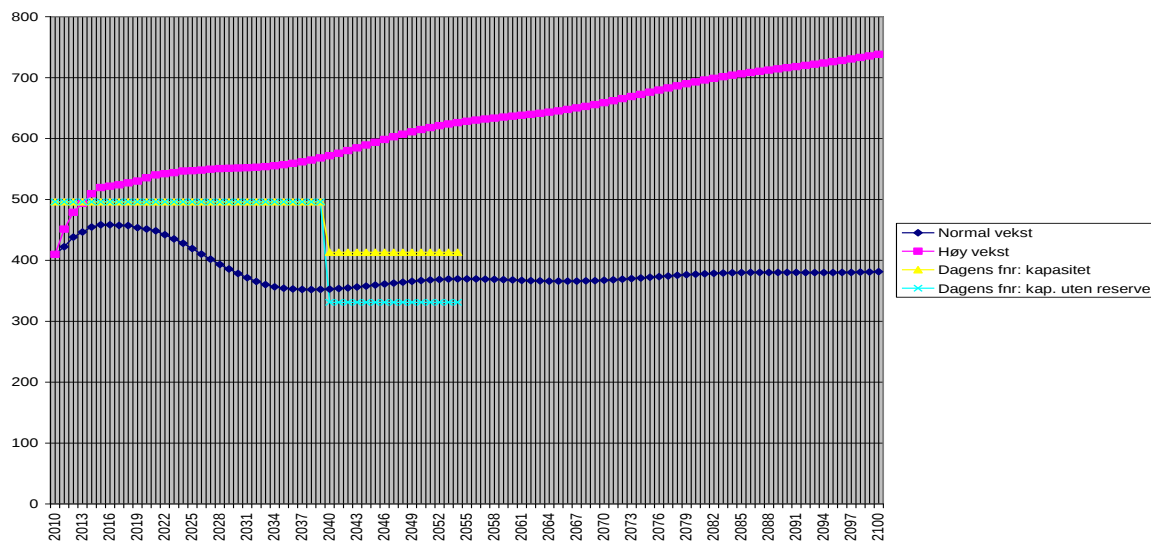
Merk at vi har modifisert scenariene i forhold til forepørselen, med en 8% vekst fra dagens nivå (lik dagens vekst) og med en «myk» overgang til et stabilt nivå for hvert av de tre scenariene. 8% tilsvarer den veksten som er observert i perioden 2004-2011 [4].



Figur 17 Scenarier for antall nye D-nummertildelinger pr. år $\tau(T)$. Scenariene løper fra 2012. Vi har også skissert faktisk tildeling av D-nummer i perioden 1991-2011, som indikerer en gjennomsnittlig årlig vekst i perioden 2004-2011 på ca. 8% (tall fra SSB [4]). Merk at faktisk tildeling av D-nummer i 2012 ser ut til å ha vært 42829 [6].

Fødselsnummer

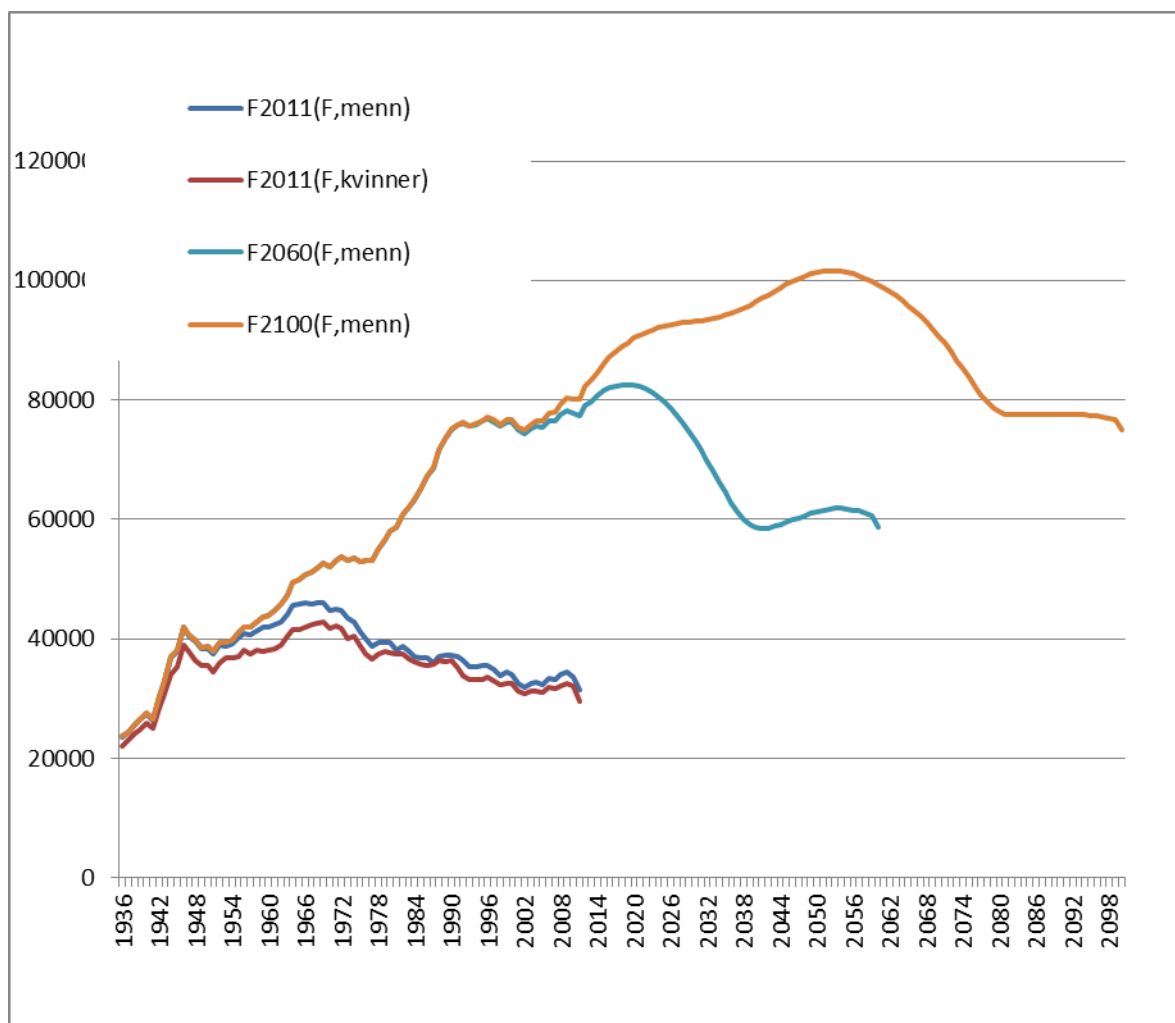
Vi har sett på scenariet for høy nasjonal vekst i forhold til tildeling av fødselsnummer, det vil si fødsler og innvandring. Dette er skissert ved følgende figur (fra [1]), som er basert på Statistisk Sentralbyrås prognoser for normal og høy befolkningsvekst frem til 2100:



Figur 18: Scenarier for nasjonal vekst: Normal vekst og høy vekst.

Legg merke til at scenariet for høy vekst bare løper frem til 2100. Vi har derfor ikke vurdert **samlete** scenarioer etter 2100.

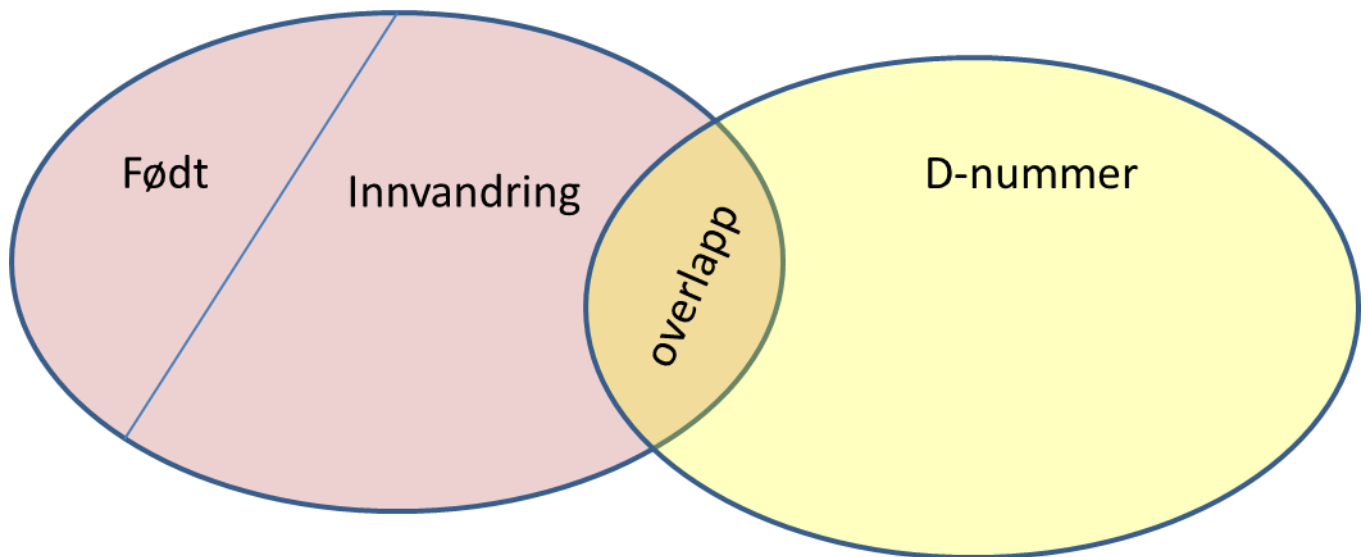
Hvis vi ser på hvordan befolkningen vil se ut med tanke på alders- og kjønnsfordeling, basert på *dagens* alders- og kjønnsfordeling samt en antagelse om at dagens alders- og kjønnsfordeling ved tildelingstidspunkt *vil holde seg konstant* også i fremtiden (se Appendiks B), så får vi følgende figur.



Figur 19 : Vekst i tildelte fødselsnummer pr årskull i henhold til vekstscenariet høy nasjonal vekst. Antagelse: Alders- og kjønnsfordeling på tildelingstidpunkt konstant og lik 2011.

Kombinert populasjon bestående av dagens fødselsnummer- og D-nummerpopulasjoner

Andel av populasjonen som fikk D-nummer i perioden 2003-2009 og som siden har fått fødselsnummer, ligger mellom 25 og 30% (tall fra SSB). For de som har fått D-nummer siden, ligger denne andelen noe lavere (ned mot 9% for de som har fått D-nummer i 2011), men det er helt klart en forsinkelse i denne prosessen og mye taler for en gjetning om at den faktiske størrelsen på «overlapp» i figuren nedenfor egentlig utgjør bortimot 30% av de som får D-nummer «nåtildags».



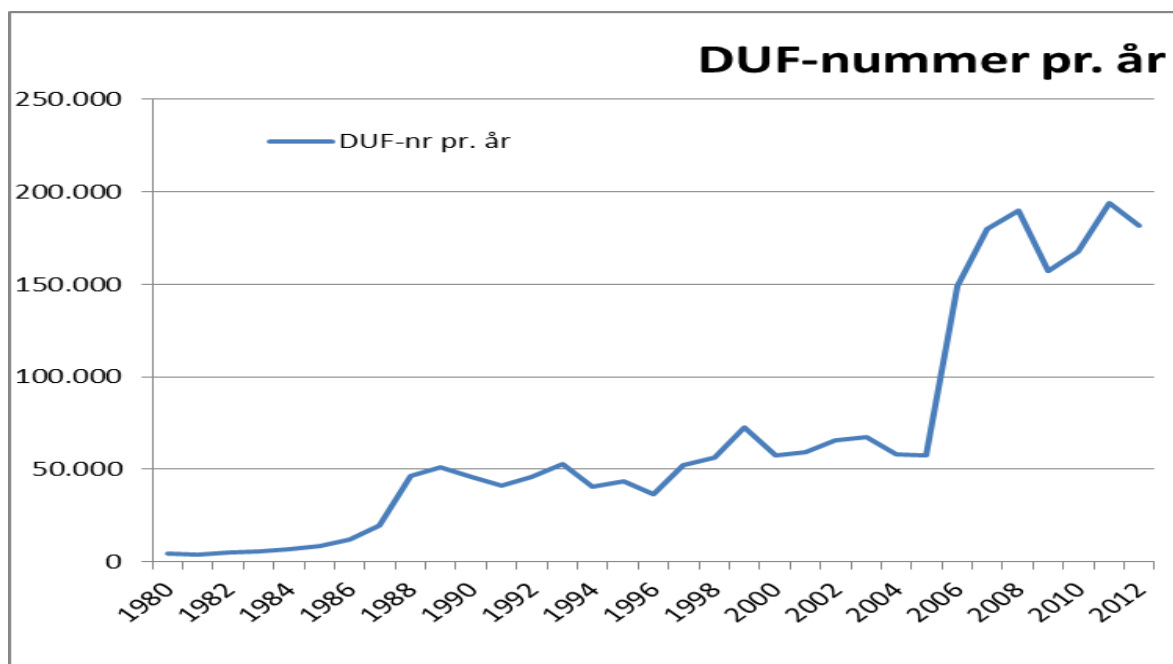
Figur 20 : Figuren viser sammenheng mellom populasjonen som får fødselsnummer og populasjonen som får D-nummer. Noen av de som får D-nummer blir senere innvandrere.

For å estimere størrelsen på og alders- og kjønnsfordelingen i den kombinerte populasjonen må vi kjenne til størrelsen på enkeltpopulasjonene og på overlappet (og vi må også vite litt mer, for alders- og kjønnsfordelingens del.) Dersom vi gjetter på et for lite overlapp, vil vi overestimere størrelsen på den kombinerte populasjonen – med andre ord tar vi «godt i» i forhold til vurderinger av behov for personidentifikator-kapasitet. Dette kan være klokt for vårt formål.

Vi har derfor i våre beregninger brukt estimatet «10% av D-nummerpopulasjone» for størrelsen på overlappet.

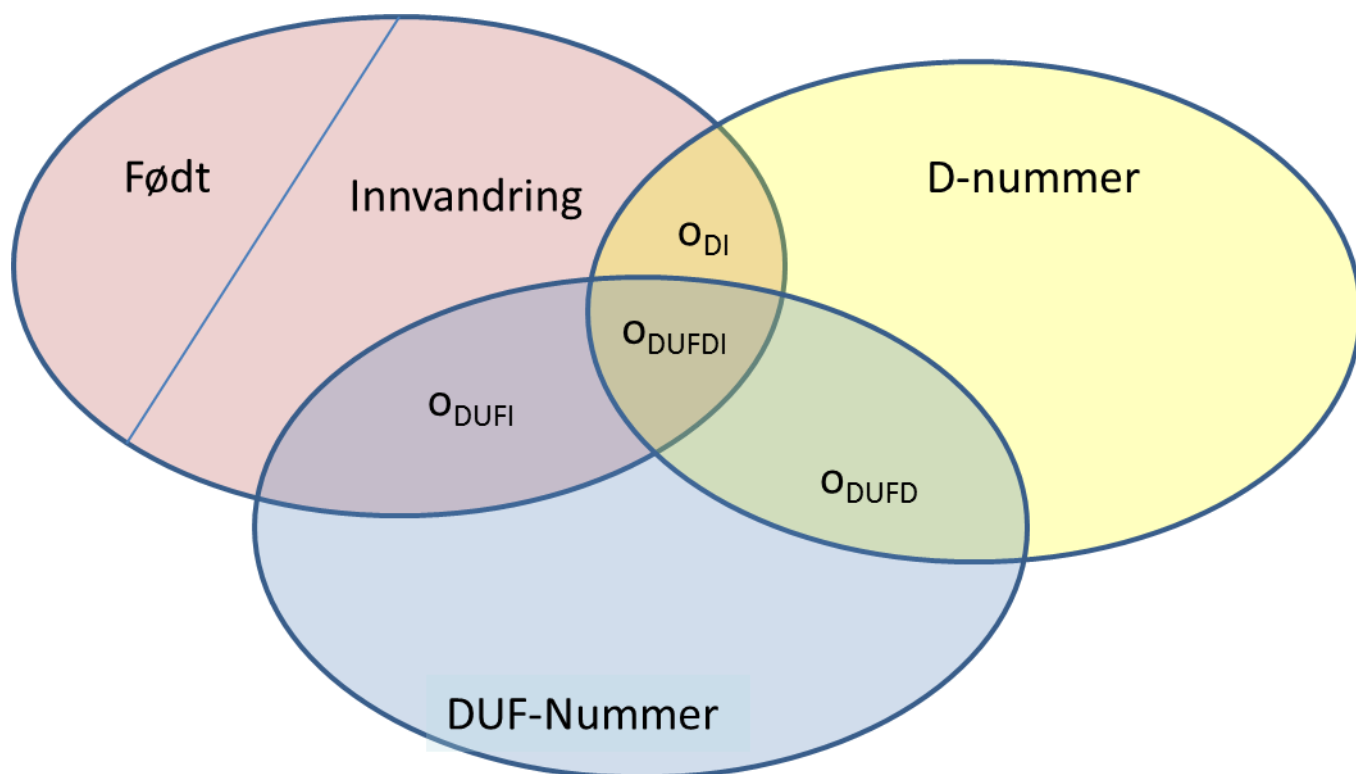
Fødselsnummer, D-nummer og DUF-nummer

Antall DUF-nummer pr. år har variert mellom 150.000 og 193.000 mellom 2006 og 2012 (dog er tallene våre fra 19.12.2012, altså ikke fullstendig,) og *viser ingen klar tendens i denne perioden.* Se figuren nedenfor.



Figur 21 Nye DUF-nummer pr. år. Kilde: Data fra SSB.

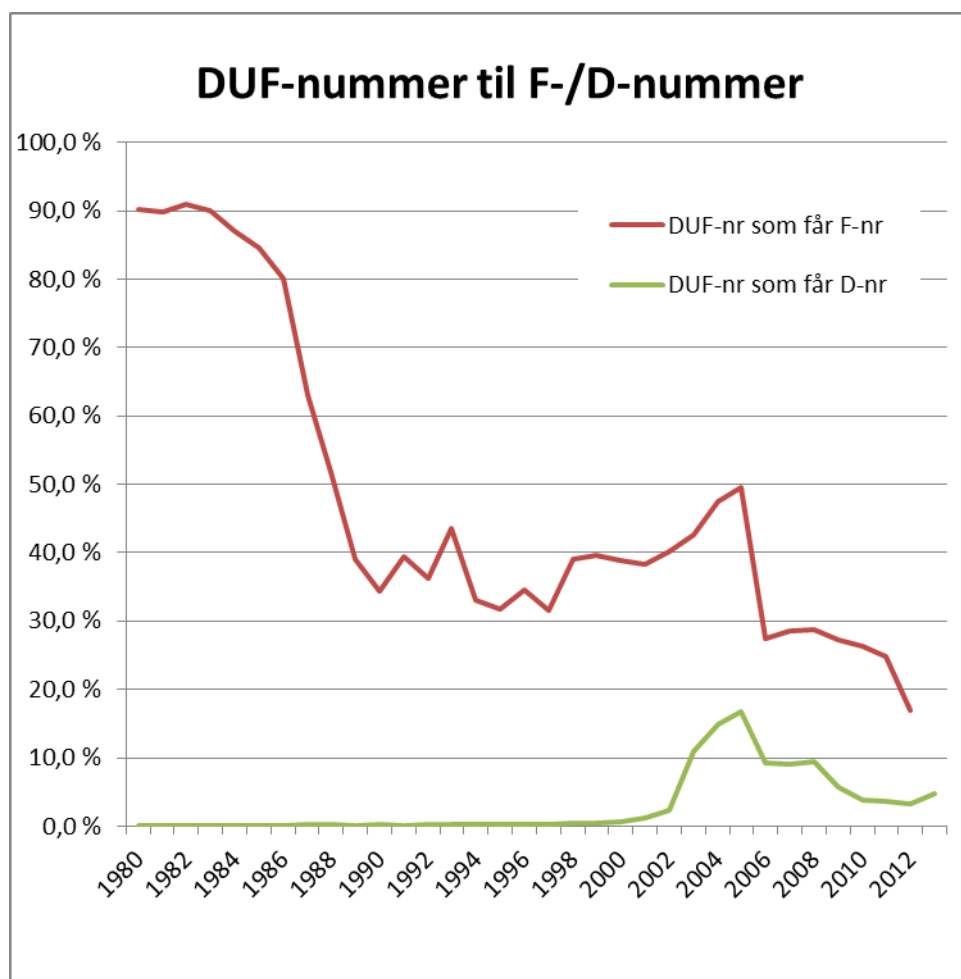
For å estimere størrelsen på en eventuelt kombinert populasjon bestående av de som i dag får enten fødselsnummer, D-nummer eller DUF-nummer, refererer vi til følgende figur:



Figur 22 Kombinert populasjon/personkrets

Det er imidlertid vanskelig å få data om hvor store overlappene er, og enda vanskeligere å finne fremtidsprognoser for disse overlappene. Når det gjelder de som har DUF-nummer og som siden får F-nummer (o_{DUF1}) gir ikke data (Figur 14) tilstrekkelig informasjon til å konkludere om det. Det kan se ut som de som har fått DUF-nummer mellom ca. 1990 og 2004 ligger på ca. 40% overgang (av populasjonen som får DUF-nummer). De som har fått DUF-nummer etter 2004 vil kanskje få F-nummer senere, selv om dette er spekulasjoner. Spekulasjonene kan imidlertid tyde på at ca. 40% av de som har DUF-nummer senere kan få status som innvandrere. Vi observerer at dette i så fall utgjør en betydelig del (mer enn halvparten) av innvandrerne.

Med hensyn til de som har DUF-nummer og som siden får D-nummer (o_{DUF2}), utgjør disse ca. 10% av DUF-nummergruppen de siste ti årene (og dermed 20% eller mer av D-nummergruppen.)



Figur 23 DUF-nummer til F-/D-nummer. Kilde: Data fra SSB.

I og med (a) at vi ikke har noe scenario å forholde oss til i forhold til vekst i DUF-nummerpopulasjonen, og (b) at vi heller ikke har grunnleggende innsikt i hvordan DUF-nummergruppen overlapper med F-nummergruppen og D-nummergruppen, og (c) at DUF-nummergruppen er svært stor i forhold til de andre gruppene, finner vi det ikke forsvarlig på denne bakgrunnen å komme med anbefalinger i forhold til kapasitet av en personidentifikator som inkluderer både F-, D-, og DUF-nummerpopulasjonene.

Helsenummer

T. Neumann [5] anslår bruken av helsenumre (H-numre) til ca. 50.000 pr. år, utenom fødsler. Men han peker også på at bruken av H-numre nettopp er aktuell enten i en akkutt-situasjon, der det ikke er en prioritert oppgave å hente inn alle pasientens personopplysninger, eller der det er vanskelig å hente opplysninger fra registrene.

Forøvrig er det vanskelig å hente prognoser for vekst i bruk av helsenumre.

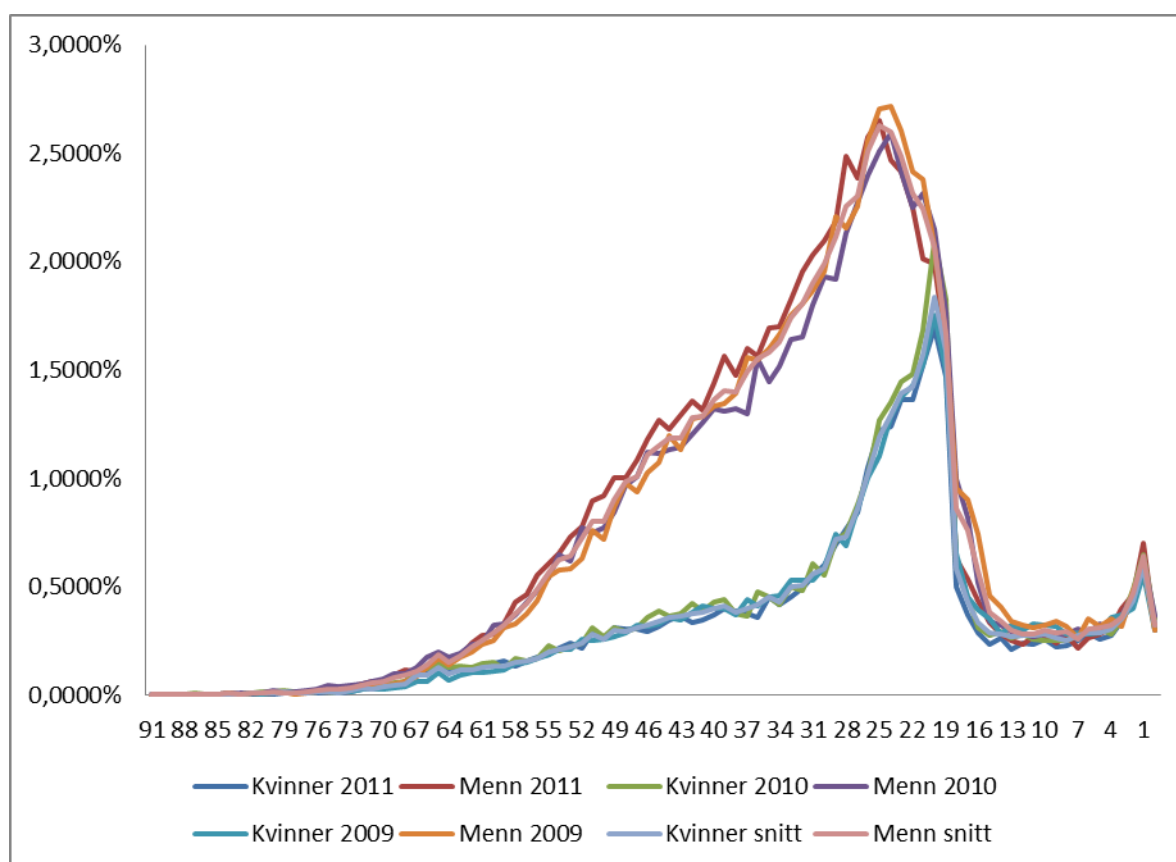
Vi anser det derfor ikke som et fornuftig administrativt tiltak å inkludere H-numre i personidentifikatoren.

Appendiks B: Aldersfordeling ved tildeling av identifikator

D-nummer

Fordelingsfunksjonen $f(X,S)$ er interessant for å anslå utviklingen av behovet for D-numre. For de nærmeste årene er valg av korrekt fordelingsfunksjon meget viktig. På lengre sikt, og dersom man har en stabil tilvekst i D-numre (slik scenariene I, II og III forutsetter), vil alle fordelingsfunksjoner gi samme resultat: $D(F,S)$ vil konvergere mot den tilsvarende kjønnsfordelte tilveksten gitt ved $\tau(I)$, som også Figur 5-7 viser.

På bakgrunn av dette gir det mening å estimere $f(X,S)$ ut fra den observerte frekvensfordeling de siste årene. Figur 24 nedenfor viser vi frekvensfordelingen observert for D-nummer tildelt i 2009, 2010 og 2011. Vi bruker **gjennomsnittet av disse** som $f(X,S)$ i våre beregninger.



Figur 12 Alders- og kjønnsfordeling for personer som blir tildelt D-nummer, for årene 2009-2011. Horisontalaksen viser alder ved tildeling. Figuren inneholder mye informasjon. For eksempel viser den at 2.6 % av alle som får tildelt D-nummer er 25 år gamle menn – dette er alders/ kjønnsgruppen som er mest vanlig. For kvinner er 20 år den mest vanlige alderen for tildeling (1.8% av alle tildelinger). **Gjennomsnittene** for menn og kvinner ligger noe høyere (henholdsvis 32 og 28 år) på grunn av den skjeve fordelingen. Merk også toppen for svært små barn.

Kommentarer:

- Vi antar at det i D-nummerpopulasjonen skjuler seg flere subpopulasjoner med hver sine karakteristiske egenskaper. Innbakt i vår matematiske $f()$ -funksjon vil det ligge forskjeller mellom vestlige og ikke-vestlige utlendinger, for eksempel. Studiet av disse de bakenforliggende

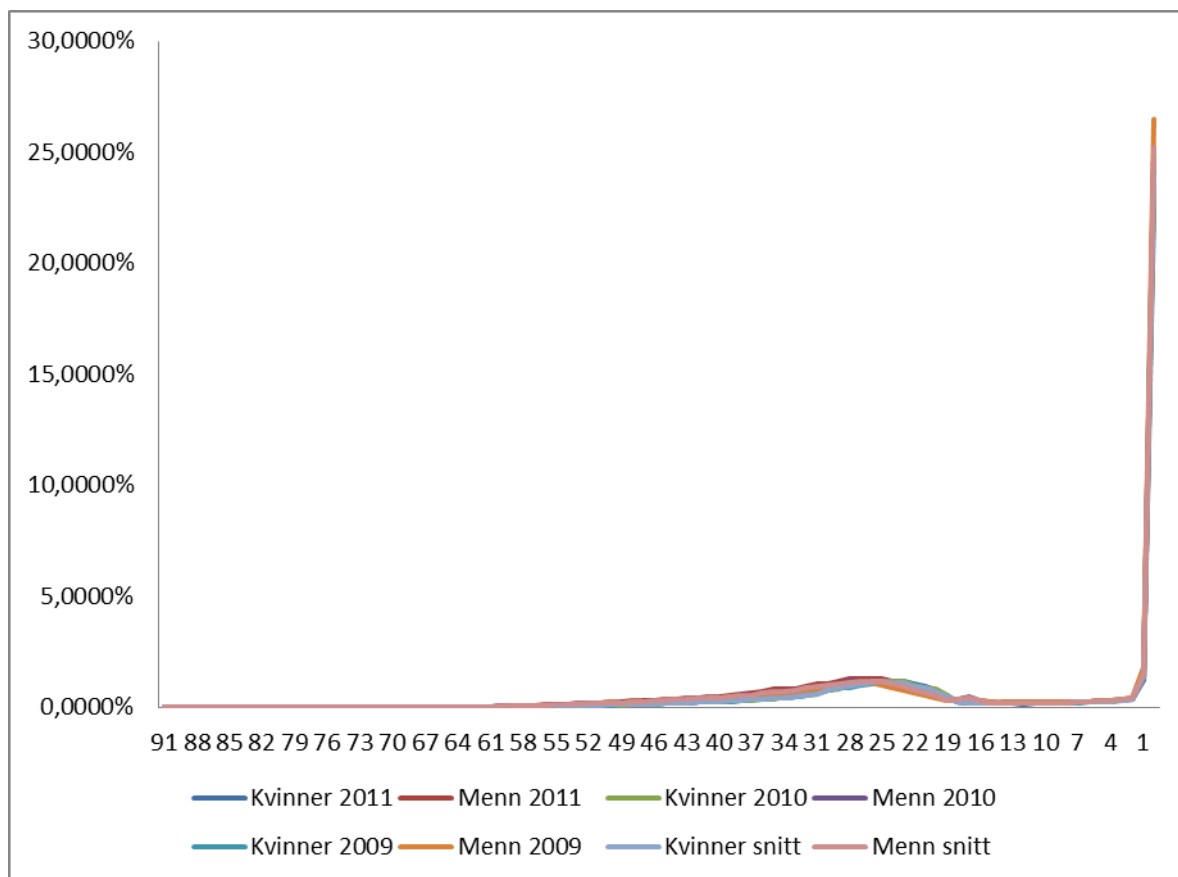
sosiologiske årsakene til disse fordelingene vil være interessant, men ligger langt utenfor vårt mandat og kompetanseområde.

- En antagelse om en $f()$ -funksjon som er konstant over tid er praktisk problematisk. Disse fordelingene er blant annet avhengig av velferdsordninger som barnetrygd og kontantstøtte.
- *Kjønnsfordeling*: Samlet kjønnsfordeling i D-nummerpopulasjonen er (pr. utgangen av 2011) ca 76% menn, ca. 24 % kvinner. Men i D-nummertildelinger 2009-2011 er fordelingen ca. 68%-32%. Forskjellen ser ut til å være at i de siste tildelingene er det mange barn, og blant disse er kjønnsfordelingen jevn.
- *Konsekvens av feil valg av $f()$ -funksjon*: Ettersom vi vurderer kapasitet pr. kjønn, er estimatet for **kjønnsfordelig** grunnleggende viktig. Vårt valg av $f()$ -funksjon tilsier 68% menn. Med hensyn til **aldersfordeling** er valg av $f()$ -funksjon *over lang tid mindre kritisk*. Som for eksempel Figur 5-7 viser, vil aldersfordelingen i befolkningen over tid flate ut på en verdi som tilsvarer årlig tilvekst, og dette er uavhengig av valg av $f()$ -funksjonen desrom tilveksten er konstant.

På kort sikt er aldersfordeling viktig. Men på kort sikt virker det også rimelig å bruke gjennomsnittsfordelingen for 2009-2011, som vi har gjort.

Fødselsnummer

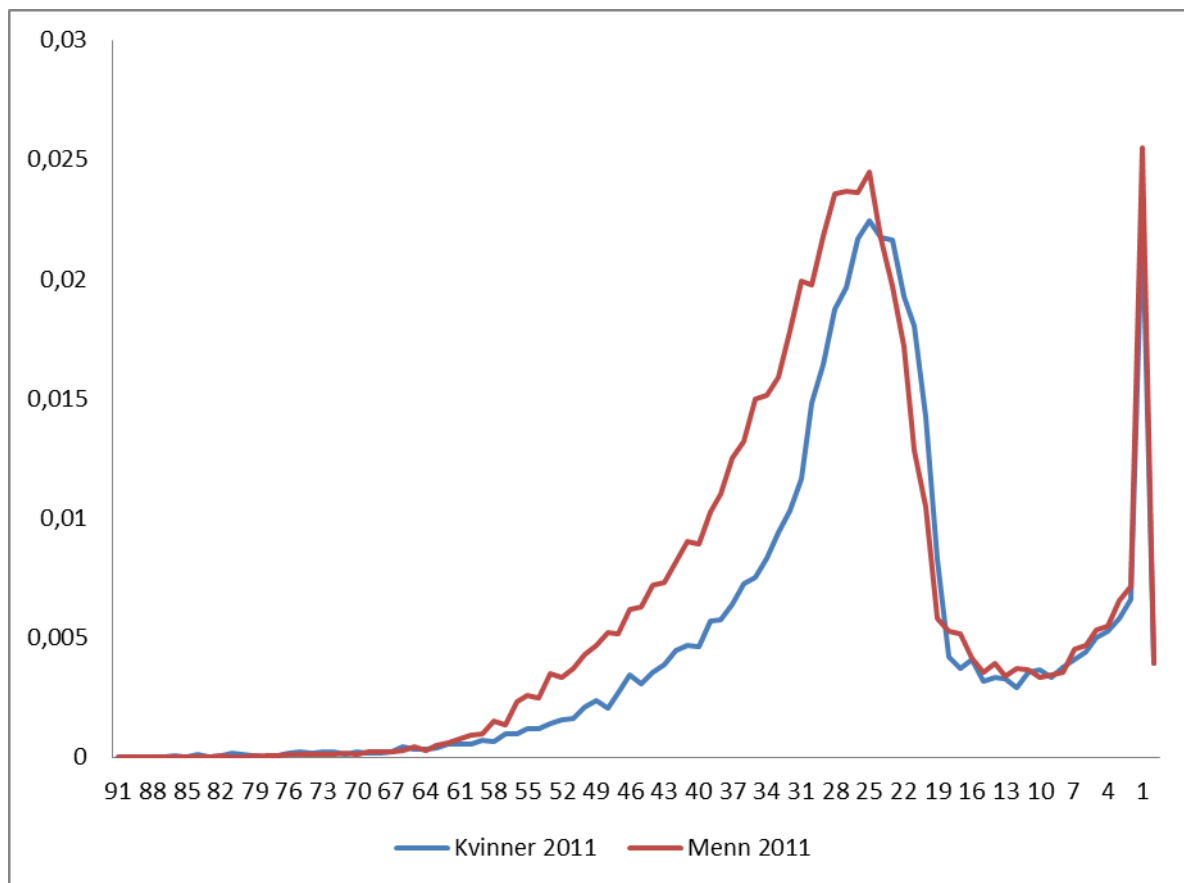
Alders- og kjønnsfordeling for personer som ble tildelt Fødselsnummer på tildelingstidspunkt, for 2011 [4], er vist i Figur 25.



Figur 13 Alders- og kjønnsfordeling for tildelte fødselsnummer. Horisontalaksen viser alder ved tildeling.

Karakteristisk for denne fordelingen er toppen for personer som er 0 år. Det kommer av at i denne populasjonen er omtrent halvparten nyfødte, det vil si ca. 25% nyfødte jenter og ca. 25% nyfødte gutter.

Dersom vi trekker ut ut gruppen av nyfødte, får vi følgende alders- og kjønnsfordeling på populasjonen av innvandrere tildelt fødselsnummer i 2011.



Figur 26 Alders- og kjønnsfordeling for tildelte fødselsnummer i 2011, der fødsler er trukket ut. Horisontalaksen viser alder ved tildeling. Figuren viser med andre ord alders- og kjønnsfordeling i innvandret populasjon.

Referanser

[1] Tor Helleseeth og Øyvind Ytrehus, «Rapport om egenskaper ved ny personidentifikator», *Rapport for Skattedirektoratet*, april 2012, revidert jan. 2013.

[2] Tor Helleseeth og Øyvind Ytrehus, « Rapport om egenskaper ved ny personidentifikator: Del 2», *Rapport for Skattedirektoratet*, september 2012, revidert jan. 2013

[3] *Forespørsel; Ny personidentifikator i Folkeregisteret: Tillegg om D-nummer og økt populasjon for personidentifikator*, Skatteetaten, jan. 2013.

[4] Data fra Kåre Vassenden, Statistisk Sentralbyrå.

[5] E-post fra Torgny Neuman til Vigdis Olsen, 12.12.12

[6] E-post fra Vigdis Olsen, 6.2.2013

Utfylt skjema sendes til vigdismerete.olsen@skatteetaten.no, tlf. 90 85 85 91

Opplysninger om utfylling:

Dato: _____

Virksomhet: _____

Avdeling: _____

Kontaktperson: _____

Telefon: _____

Epostadresse: _____

Samfunnsøkonomisk analyse av å endre dagens fødselsnummer – Spørreskjema

Skattedirektoratet (SKD) har sammen med Finansdepartementet bestemt å gjennomføre en konseptvalgutredning (KVU) av å innføre ny personidentifikator (fødselsnummer) i Folkeregisteret. Konseptvalgutredningen skal gi anbefaling om valg av ny identifikator på bakgrunn av en samfunnsøkonomisk analyse. Dette spørreskjemaet med veiledning sendes til utvalgte virksomheter for å samle inn data om konsekvensene, dvs. nytte- og kostnadsvirkninger av å endre dagens fødselsnummer.

Erfaring har vist at det kan være hensiktsmessig å ha et intervju/samtale for å følge opp og utdype en del av svarene i skjemaet. Det kan også være behov for å forklare skjemaet underveis i prosessen. Vi ber om at dere med utgangspunkt i denne veilederen identifiserer mulige konsekvenser av å innføre ny personidentifikator i egen virksomhet, og at dette kan være utgangspunkt for en mulig oppfølgingssamtale.

Konsekvensene av en ny personidentifikator er komplekse og omfattende, og vår tilnærming er at dataene samles inn gjennom en prosess som består av følgende steg:

1. Virksomheten identifiserer arbeidsprosesser, IKT-system og type konsekvenser. Dere svarer på spørsmål i skjemaet så langt det lar seg gjøre.
2. Ved behov gjennomføres intervju/samtale med SKD/prosjektet med tid for spørsmål, ideer og konkrete innspill. Spørreskjemaet blir gjennomgått hvor virksomhetene går igjennom sine første innspill, drøfting og avklaringer.
3. Eventuelt etterarbeid i virksomheten. Arbeidet med å beskrive konsekvenser ferdigstilles, både prissatte og ikke-prissatte virkninger skal konkretiseres. Utfylt skjema for virksomheten sendes SKD ved prosjektet.

Spørreskjemaet har følgende oppbygging:

1. Kort om prosjektet
2. Fødselsnummerets oppbygging
3. Bruk av dagens fødselsnummer
4. Mulige overgangsordninger
5. Alternative personidentifikatorer
6. Overgangsordninger knyttet til personidentifikator
7. Konkretisering av prissatte virkninger (inkl. vedlagt regneark)
8. Konkretisering av ikke-prissatte virkninger.
9. Oppsummering

Til de fleste spørsmål er det en kort omtale av tema og forklaring av selve spørsmålet.

1. Kort om prosjektet

Dagens personidentifikator i Folkeregisteret, fødselsnummeret, har grunnet kapasitetshensyn en antatt maksimal levetid frem til 2039. På denne bakgrunn er det behov for å endre dagens identifikator. Finansdepartementet har gitt Skattedirektoratet i oppdrag å utrede nærmere de samfunnsmessige, økonomiske og administrative konsekvenser knyttet til valg av ny identifikator. Prosjektet sees i sammenheng med strategi for nytt Folkeregister og inngår i Skattedirektoratets program for modernisering av Folkeregisteret.

Finansdepartementet og Skattedirektoratet har bestemt at det gjennomføres en konseptvalgutredning (KVU) i henhold til Finansdepartementets retningslinjer og rammeverk. Hovedmålet for prosjektet er å utrede de samfunnsøkonomiske konsekvensene knyttet til ny personidentifikator i et slikt omfang at det kan benyttes som beslutningsgrunnlag for valg av alternativ.

Dette innebærer at det skal gjennomføres en samfunnsøkonomisk analyse med relevante alternative personidentifikatorer. De samlede nytte- og kostnadsvirkninger skal identifiseres, beskrives og kvantifiseres så langt det er forsvarlig. Dette skal gjøres gjennom beskrivelse av konsekvenser og representative estimer for berørte virksomheter.

2. Fødselsnummerets oppbygging

Dagens fødselsnummer har følgende oppbygging:

D¹D²M¹M²Å¹Å²I¹I²I³K¹K²

D: dato, M: måned, Å: år, I: individsiffer/personnummer, K: kontrollsiffer. I³ viser kvinne (like tall) eller mann (ulike tall).

Fortsatt bruk av dagens fødselsnummer vil inngå i analysen som en referanse som de andre alternativene måles opp i mot (nullalternativet).

3. Bruk av dagens fødselsnummer

3.1. Innledning

Nedenfor har vi beskrevet noen generelle bruksområder hvor fødselsnummeret benyttes. Listen nedenfor er kun til illustrasjon og er ikke en komplett liste. Hver virksomhet er forskjellig, og fødselsnummerets bruksområder vil variere. Noen eksempler på bruksområder for fødselsnummeret (dagens personidentifikator) er:

- Registrering av person, kunde, klient, pasient i databaser/registre
- Registrering av eierskap/tilknytning til eiendom, kjøretøy, løsøre, foretak etc.
- For entydig identifisering av personer
- Finne opplysninger om individer både fra Folkeregisteret og egne systemer/databaser. Kan være i forbindelse med kundebehandling, saksbehandling, pasientbehandling etc.
- Utveksling av personopplysninger mellom virksomheter og internt i en egen virksomhet
- Pålagt rapportering til offentlige myndigheter
- Ved utstedelse av legitimasjonspapirer (bankkort, vernepliktsbok, sertifikat, helsekort, pass, studiebevis, osv.)
- Beregne personers alder og opplysninger om kjønn. Fødselsnummeret gir enkel tilgang til dette

3.2 Spørsmål

Nedenfor følger tre spørsmål om virksomhetens bruk av dagens fødselsnummer.

3.2.1. Beskriv kort og overordnet hvilke arbeidsområder/hovedprosesser hvor virksomheten i dag benytter fødselsnummer?

Svar:

3.2.2. Hvor mange og hvilke typer datasystemer vil bli berørt når fødselsnummeret endres?

Svar:

3.2.3. Hvem er de viktigste virksomhetene/aktører dere har datautveksling med basert på fødselsnummer?

Svar:

4. Mulige overgangsordninger

En ny identifikator med tilstrekkelig kapasitet kan utformes på mange ulike måter. For eksempel kan lengden endres, informasjonsbærende elementer kan tas ut og kontrollsiffer kan fjernes. Vi kommer tilbake til detaljer mht. utforming senere i spørreskjemaet. Først vil vi fokusere på overgangsordningen, og dens betydning for konsekvensene av å innføre en ny identifikator.

Overgangsordninger beskriver prosessen med å innføre en ny identifikator, dvs. hvordan man skal komme fra dagens situasjon og til en situasjon hvor innføring av ny personidentifikator er sluttført.

Hensikten med overgangsordningen er å minimere kostnader og ulemper for virksomhetene. Dette er en periode hvor virksomhetene kan planlegge, utvikle og implementere ny(e) løsning(er).

Når vi vurderer overgangsordninger er følgende tre dimensjoner viktig:

A. Hvor lang tid virksomhetene har til innfasing av ny ordning,

B. Hvilke grupper som skal tildeles ny identifikator.

C. Hvilke tekniske tjenester folkeregistermyndigheten kan tilby for å lette overgangen

A. Tid til innfasing

Hvis vi tar utgangspunkt i at ny identifikator besluttet innen 2015, vil det være maksimum 25 år fra beslutning til ny identifikator må være innført i år 2040. Dette er det lengste tidsløpet virksomhetene har til å planlegge, utvikle og implementere nye løsninger hvor ny personidentifikator benyttes. I motsatt ende kan vi forutsette at 5 år er den korteste tidsperioden, dvs. at virksomhetene blir gitt 5 år til å utvikle løsninger for og fase inn en ny personidentifikator.

Det er opp til den enkelte virksomhet å bestemme hvordan egen innfasing skal skje innenfor det definerte tidsrom. For eksempel hvor lang periode man velger til planlegging, utvikling og implementering av nye løsninger før ordinær drift.

En lang overgangsperiode kan gi mulighet for en naturlig oppgradering/utskifting av dagens IKT-systemer hvor det inkluderes løsning for bruk av ny personidentifikator. Man har med andre ord kjennskap til hva som må endres, og kan gjennomføre dette uten å foreta ekstraordinære endringer, men inkludere dette inn i en naturlig oppgradering/utskifting. Omfanget og kostnaden for å kunne ta i bruk ny løsning med ny personidentifikator vil kunne avhenge av hvilken identifikator som velges. Besvar derfor nedenstående fem spørsmål overordnet og på et generelt nivå.

4.a. Spørsmål

Gitt en overgangsperiode på 10 år:

4.a.1. Gjør et anslag over hvor stor andel av virksomhetens IKT-systemer som benytter fødselsnummeret (ref. pkt. 3.2.2), hvor en løsning for ny personidentifikator kan bli endret/skiftet ut gjennom en naturlig oppgradering/utskifting i løpet av 10 år (f.eks. nye versjoner, nye moduler, nyutvikling etc.)?

Svar:

4.a.2. Vil planlagte oppgraderinger/utskiftninger av eksisterende IKT-systemer i løpet av 10 år kunne redusere kostnadene ved å innføre en ny personidentifikator i forhold til å måtte gjøre dette over en 5 års periode? Beskriv kort hvordan dette i tilfelle kan skje?

Svar:

Gitt en overgangsperiode på 20 år:

4.a.3. Gjør et anslag over hvor stor andel av virksomhetens IKT-systemer som benytter fødselsnummeret (ref. pkt. 3.2.2), hvor en løsning for ny personidentifikator kan bli endret/skiftet ut gjennom en naturlig oppgradering/utskifting i løpet av 20 år (f.eks. nye versjoner, nye moduler, nyutvikling etc.)?

Svar:

4.a.4. Vil planlagte oppgraderinger/utskiftninger av eksisterende IKT-systemer i løpet av 20 år kunne redusere kostnadene ved å innføre en ny personidentifikator i forhold til å måtte gjøre dette over en 5 års periode? Beskriv kort hvordan dette i tilfelle kan skje?

Svar:

Andre forhold

4.a.5. Er det andre forhold knyttet til overgangsordninger som dere mener bør vektlegges?

Svar:

B. Tildeling av ny identifikator

Et viktig spørsmål er om hele befolkningen skal få ny identifikator eller om kun nye grupper får ny identifikator.

Dersom alle grupper i befolkningen skal tildeles ny identifikator, vil dette være en stor omstilling som krever skifte av alle eksisterende fødselsnummer, utstedelse av nye identitetspapirer etc.

Alternativt kan de som har fått tildelt fødselsnummer etter dagens ordning, beholde dette livet ut.

4.b Spørsmål

4.b.1. Hvilke utfordringer ser virksomheten ved å operere med to forskjellige personidentifikatorer (ny og gammel) i parallell?

Svar:

4.b.2. Hvilke utfordringer vil det gi dersom personer skal skifte fra gammel til ny personidentifikator? Bør dette tas som en big bang eller gjennom en nærmere angitt tidsperiode? Hvor lang bør i så fall tidsperioden være?

Svar:

4.b.3. Hvilke utfordringer knytter seg til bruk av identitetsdokumenter eller andre papirer/dokumenter, kort etc. knyttet til personer med et skifte av personidentifikator?

Svar:

C. Tekniske tjenester

For enkelte personer er flere identifikatorer i bruk. Dette kan gjelde utlendinger som i første omgang fikk D-nummer, men som senere har bosatt seg og fått fødselsnummer. Tilsvarende har enkelte personer endret fødselsnummer grunnet feilregistrering eller ny dokumentasjon av fødselsdato. Disse personene har i dag fått tildelt to identifikatorer, en som er utgått og en ny.

Folkeregistermyndigheten tilbyr i dag tjenester som returnerer/viser nytt fødselsnummer ved søk på utgått nummer. Det finnes også abonnementstjenester for ajourhold av egne registre. Hvis eksisterende populasjon skal tildeles ny identifikator, ser man for seg at det som en del av overgangsordningen, tilbys tilsvarende oversettelsestjenester. Dette er aktuelt for alternativ 4, informasjonsløs og alfanumerisk identifikator med 8 posisjoner (ref. kap. 5)

4.c Spørsmål

4.c.1. Hvilke positive eller negative effekter vil bruken av slike «oversettings-tjenester» ha for virksomheten?

Svar:

4.c.2. Hvordan vil tilgjengeligheten / bruken av slike tekniske «oversettings -tjenester» påvirke svaret på spørsmål 4.b.1 (Hvilke utfordringer ser virksomheten ved å operere med to forskjellige personidentifikatorer (ny og utgått) i parallell)?

Svar:

4.c.3. Ser dere behov for andre tekniske tjenester, i tillegg til slike «oversettings-tjenester»? I så fall hvilke?

Svar:

5. Alternative personidentifikatorer

Det er identifisert et stort antall mulige nye personidentifikatorer. Gjennom grovsiling er det søkt å få antallet ned til et håndterbart utvalg som både inkluderer de mest aktuelle alternativene som tilfredsstillende absolutte krav til kapasitet og samtidig viser tilstrekkelig av spennet i mulighetsrommet, jfr. krav til KVV. Det reduserte antall personidentifikatorer legges til grunn for den videre kartleggingen av konsekvenser.

I tabellen nedenfor har vi tatt utgangspunkt i to dimensjoner ved en ny personidentifikator, slik at det blir fire hovedtyper av identifikatorer.

Den første dimensjonen er tegntype. Numeriske tegn er en rekke med tall. Alfnumeriske tegn er en rekke med tall og bokstaver i kombinasjon.

Den andre dimensjonen omhandler informasjon, dvs. om identifikatoren er informasjonsbærende eller informasjonsløs. Dagens fødselsnummer gir informasjon både om fødselsdato og kjønn. Dersom det er en tilfeldig rekke av tall eller bokstaver, vil identifikatoren være informasjonsløs.

I tabellen nedenfor er det i tillegg til dagens fødselsnummer angitt fire aktuelle personidentifikatorer. Det er disse alternative identifikatorene som er utgangspunkt for spørsmålene videre i skjemaet.

Informasjon Tegntype	Informasjonbærende	Informasjonsløs
Numeriske tegn	0. Dagens fødselsnummer, to kontrollsifre (11): DDMMÅÅIIKK 1. Ett kontrollsiffer ellers samme oppbygning som alt. 0 (11): DDMMÅÅIIK	3. Kundenummer med to kontrollsifre (11): IIIIIIIIKK Har to varianter 3a og 3b
Alfanumeriske tegn	2. Individnummer og kontrollsifre alfanumerisk (11). To kontrollsifre. Første bokstav angir århundre: DDMMÅÅAAKK	4. Informasjonsløs og alfanumerisk (8). To kontrollsifre: AAAAAAKK

DD=dag, MM=måned, ÅÅ=år, I=individnummer, A=alfanumerisk tegn, K=kontrollsiffer

Nedenfor er alternativene i tabellen ovenfor forklart:

0. Dagens fødselsnummer - 11 posisjoner, 2 kontrollsiffer
 Dag, måned og år oppgitt med til sammen seks posisjoner. Deretter kommer et individnummer på tre sifre hvorav siste siffer angir kjønn, kvinner har like tall og menn ulike tall. To kontrollsifre som beregnes etter en såkalt 0 modulo 11 kontroll.
1. Dagens fødselsnummer med ett kontrollsiffer - 11 posisjoner.
 Grunndatarapportens subsidiære forslag. Det frigjorte kontrollsymbolet K går inn i individnummeret som dermed blir på fire sifre. Kan benytte samme feilkontroll som i dag (modulo 11 kontroll). Alternativet skilles fra dagens nummerserie ved å benytte S modulo 11 kontroll der S er kontrollsummen som i dag bestemmer det frigjorte kontrollsifferet K. I dagens fødselsnummer er K bestemt slik at S er lik 0 modulo 11. I det nye systemet kan S anta verdiene fra 1 til 10. Hver verdi av S fra 1 til 10 har en kapasitet på 826 per dato. Ved å tildele identifikator kun innen gitte verdier av S for gitte tidsperioder, kan det skilles på århundre. F.eks. kan det for S fram til 2099 benyttes verdiene 1 og eventuelt 2 (hvis det skulle bli behov over 826 identifikasjoner per dato). Fra og med år 2100 benyttes påfølgende verdier osv. Tildelte fødselsnummer etter dagens ordning kan beholdes, dvs. at personer som har fødselsnummer når ny identifikator trår i kraft, beholder sitt gamle. Det anbefales at alternativet kombineres med nytt felt i Folkeregisteret hvor århundre for fødselsdato angis, dette kan beregnes automatisk ut ifra fødselsnummeret, både for ny og gammel identifikator.
2. Individnummer og kontrollsifre er alfanumeriske. Alternativet har to kontrollsifre og 11 posisjoner totalt. Tildelte fødselsnummer etter dagens ordning kan beholdes, dvs. at personer som har fødselsnummer når ny identifikator trår i kraft, beholder sitt gamle. Blant de 5 posisjonene for individnummer og kontrollsifre må minst tre posisjoner være bokstaver. Kravet er satt for å skille ny personidentifikator fra dagens fødselsnummer samt oppdage feilskrivning etc. Det sjekkes først om det er gammel eller ny identifikator (om det er minst tre bokstaver eller ikke). Hvis det er identifikator fra dagens fødselsnummer, benyttes eksisterende rutine for feilkontroll. I motsatt fall benyttes ny rutine for feilkontroll som anbefales å være modulo 23 kontroll. Århundre skilles gjennom første bokstav. Systemet har

en kapasitet på minst 7535 identitetsnumre per dato (DDMMÅÅ), eller ca. $7535/13 = 580$ for hver bokstav-verdi. Det benyttes en bokstav for hvert kjønn fram til og med 2099. Fra og med 2100 benyttes de neste to bokstavene osv.

3. Informasjonsløst kundenummer med 11 sifre hvorav 2 er kontrollsifre. Alternativet kan karakteriseres som et informasjonsløst kundenummer som tildeles tilfeldig. Det betyr at man ikke tildeler kontinuerlig f.eks. i en stigende eller synkende nummerserie, da det vil kunne gi opplysninger om personens fødselsdato, tidspunkt for registrering etc. Vi har her to varianter 3a og 3b:

3a: Tildelte fødselsnumre etter dagens ordning kan beholdes, dvs. at personer som har fødselsnummer når ny identifikator trår i kraft, beholder sitt gamle.

3b: Det fattes et "prinsippvedtak" om at alle skal over på informasjonsløs personidentifikator, noe som innebærer at også de som har fødselsnummer etter dagens ordning skal få ny informasjonsløs identifikator.

Man må unngå å tildele allerede benyttede nummerserier. Det kan benyttes samme feilkontroll (modulo 11 kontroll) som ved dagens fødselsnummer.

4. Informasjonsløst alfanumerisk kundenummer med totalt 8 posisjoner hvorav to kontrollsifre. Det benyttes tilfeldig tildeling for ikke å angi opplysninger om fødselsdato, tidspunkt for registrering etc. Det benyttes ny rutine for feilkontroll som anbefales å være modulo 23 kontroll. Alternativet krever at personer med eksisterende fødselsnummer, blir tildelt ny identifikator etter nytt regime (informasjonsløst alfanumerisk med 8 posisjoner).

5.1 Spørsmål

Nedenfor er det noen spørsmål knyttet til de alternative personidentifikatorene, gi en overordnet beskrivelse.

5.1.1. Er det viktig at identifikatoren har nøyaktig 11 posisjoner? Hvilke utfordringer oppstår om det er færre eller flere enn 11 posisjoner (gi gjerne eksempler)?

Svar:

5.1.2. Vil det medføre ekstra kostnader å anvende alfanumeriske tegn versus numeriske tegn?

Svar:

5.1.3. Vil det medføre ekstra kostnader å gå fra to til ett kontrollnummer? Hva hvis moduluskontrollen endres?

Svar:

5.1.4. Benytter virksomheten dagens fødselsnummer som informasjon om alder (dag, måned og år) og kjønn?

Svar:

5.1.5. Hvis informasjon om alder brukes, hvilke av elementene dag, måned og år brukes?

Svar:

5.1.6. Vil det være behov for å legge til nye elementer i virksomhetens egne IKT-systemene dersom man skal anvende informasjonsløs identifikator?

Svar:

6. Overgangsordninger knyttet til personidentifikator

For å kunne beskrive konsekvensene av å innføre ny personidentifikator er det behov for å klargjøre alternative overgangsordninger mer konkret. Hensikten med overgangsordninger er å gjennomføre endringsprosessen slik at omstillingskostnadene blir lavest mulig, samtidig som interessentene tas hensyn til. Det vil si en ordning som fortrinnsvis både er akseptabel, håndterbar og økonomisk forsvarlig.

For de ulike alternative personidentifikatorer er følgende forhold ved overgangsordninger aktuelle:

- Hvis personer med gammel identifikator skal få ny identifikator, må det vurderes om ID-dokumenter skal fornyes/endres. Det gjelder også andre id-dokumenter enn pass samt kort og formelle papirer knyttet til person hvor dagens fødselsnummer benyttes.

- Skal alle få ny identifikator over natten (big bang) eller skal det skje en gradvis og smidig innføring. For eksempel kan ny identifikator vurderes tildelt samtidig med at pass fornyes.
- Pris for endring av tekniske løsninger – tilrettelegging av IKT-systemer

6.a. Spørsmål

I kapittel 4 ble mulige overgangsordninger behandlet, mens det i kapittel 5 ble gjennomgått alternative personidentifikatorer. Ut i fra hvilken personidentifikator man velger, vil det kunne være behov for forskjellige overgangsordninger samtidig som samme overgangsordning kan ha forskjellig karakter/innhold.

Nedenfor er det laget en tabell med et felt for hver alternativ personidentifikator (ref. kapittel 5). Trolig vil besvarelsen fra en identifikator delvis kunne gjenbrukes for andre identifikatorer. I slike tilfeller er det tilstrekkelig å gjøre en henvisning. Beskriv kort i tabellen for alternativene 1-4 hva som særtegnert eller er viktig knyttet til overgangsordninger for den respektive personidentifikatoren.

Identifikator	Stikkord for overgangsordning
Alternativ 0 Dagens fødselsnr.	<i>Ikke behov for noen overgangsordning.</i>
Alternativ 1 Informasjonsbær. numerisk 11 posisjoner, 1K	
Alternativ 2 Informasjonsbær. alfanumerisk 11 posisjoner, 2K	
Alternativ 3a Informasjonsløs numerisk 11 posisjoner, 2K	
Alternativ 3b Som 3a, men hvor alle skifter til informasjons- løs identifikator. Numerisk, 11 posisjoner, 2K	
Alternativ 4 Informasjonsløs alfanumerisk 8 posisjoner, 2K	

7. Konkretisering av prissatte virkninger

Aktuelle kostnader kan være innenfor områdene:

- Investering: Planlegging, systemendring/utvikling, test og produksjonssetting etc.
- Forvaltning, drift og vedlikehold (FDV)

Det er vedlagt et regneark som viser et eksempel på oppstilling av prissatte virkninger. Her er listet opp noen få kostnadsposter som kan fylles ut for aktuelle system eller prosjekt hvor ny personidentifikator vil få betydning (kolonnene D-K).

Ved behov utvides antall kolonner. Erstatt kolonnenavn System 1, System 2 osv. med riktig navn.

Kolonne B er summen av utfylte kolonner for de enkelte system. Virksomheten står fritt i å gjøre endringer i regnearket eller benytte en annen oppstilling/struktur.

Regneark som viser virksomhetens estimering, kan sendes sammen med utfylt spørreskjema.

Kostnader eller gevinster/innsparinger skal beregnes som differansen opp mot en situasjon der dagens fødselsnummer videreføres (nullalternativet).

7.1. Spørsmål

Besvar overordnet på følgende:

7.1.1 Identifiser typer prissatte kostnader virksomheten vil ha ved innføring av ny personidentifikator.

Svar:

7.1.2: Estimer kostnadene angitt på punktet foran (7.1.1) for de 4 alternative personidentifikatorene (tallgrunnlaget kan hentes fra vedlagte regneark).

Svar:

7.1.3. Identifiser om virksomheten vil ha gevinster av å endre dagens fødselsnummer til de aktuelle alternativene skissert under kapittel 5.

Svar:

8. Konkretisering av ikke-prissatte virkninger

Nytte- og kostnadsvirkningene skal kvantifiseres så langt det er faglig forsvarlig. De virkningene som ikke er behandlet i kapittel 7, beskrives og vurderes i dette kapittelet.

Konsekvensene for ikke-prissatte effekter vurderes gjennom betydning og omfang for hver effekt og alternativ. Det benyttes en 9 trinns konsekvensskala fra meget stor positiv (++++) til meget stor negativ konsekvens (- - - -). Skalaen har følgende kvalitative verdier:

++++	Meget stor positiv
+++	Stor positiv
++	Middels positiv
+	Liten positiv
0	Ubetydelig
-	Liten negativ
--	Middels negativ
---	Stor negativ
----	Meget stor negativ

Nedenfor er det laget en tabell hvor de enkelte ikke-prissatte virkninger kan vurderes ved å angi konsekvens (iht. 9 trinns skalaen). Vurdering av alternativ 1-4 skal gjøres i forhold til nullalternativet. Ta med de viktigste virkningene.

Ikke-prissatt virkninger	Konsekvens				
	Alt. 1	Alt. 2	Alt. 3a	Alt. 3b	Alt. 4

Ikke-prissatt virkninger	Konsekvens				
	Alt. 1	Alt. 2	Alt. 3a	Alt. 3b	Alt. 4

Angi nedenfor gjerne forklarende tekst knyttet til den enkelte ikke-prissatt virkning:

Kommentarer:

9. Oppsummering

Nedenfor er det ønskelig at virksomheten kunne komme med noen oppsummerende hovedkommentarer eller synspunkter:

Kostnadsposter	Kostnader oppgis i kroner
----------------	------------------------------

Kostnader
oppgis i kroner

Systemer, prosjekter* etc. som blir berørt av ny personidentifikator *Hvis nødvendig

*Fylles ut etter behov. Erstatt System 1 osv. med riktig navn lag flere system

*Hvis nødvendig
lag flere system

[illegible]

Alternativ 2 (DDMMÅAAAKK)	Sum kr.		System 1	System 2	System 3	System 4	System 5	System 6	System 7	System 8*
Planlegging	0									
Systemendringer/utvikling (inkl. konvertering)	0									
Test og produksjonssetting	0									
Andre kostnader	0									
SUM investering	0		0	0	0	0	0	0	0	0
ÅRLIG endring i forvaltning, drift og vedlikehold (FDV)	0									

Alternativ 3a (IIIIIIKK)	Sum kr.		System 1	System 2	System 3	System 4	System 5	System 6	System 7	System 8*
Planlegging	0									
Systemendringer/utvikling (inkl. konvertering)	0									
Test og produksjonssetting	0									
Andre kostnader	0									
SUM investering	0		0	0	0	0	0	0	0	0
ÅRLIG endring i forvaltning, drift og vedlikehold (FDV)	0									

[illegible][illegible]

MENNESKELIGE FAKTORER VED VALG AV NY PERSONIDIKATOR

Forfatter:	Laura Arlov og Gautam Ghosh
Dato:	10.01.2012
Sist endret:	04.04.2013
Dok.referanse:	brukskvalitesaspekter av pid.doc
Rev./Ver.:	0.3
Status:	utkast

Dokumenthåndtering

Endringslogg

Dato	Forfatter	Versjon	Endring
05.01.2012	Laura Arlov	0	Opprettet dokumentet
24.02.2012	Laura Arlov	0.2	Utkast til innhold
05.12.2012	Gautam Ghosh	0.3	Lagt til avsnitt (Knut Nordby sitat)

Deltakere

Navn	Rolle
Laura Arlov	Forfatter
Gautam Ghosh	Forfatter

Distribusjon

Navn	Organisasjon

Innhold

Dokumenthåndtering.....	ii
1 Innledning	4
Hva er Personindikator (PID)?	4
2 Hvem er sluttbrukere?	5
Ulike sluttbrukergrupper.....	5
3 Brukerbehov	7
3.1 Organisasjoners behov.....	7
3.2 Individets behov.....	7
Behov 1: Si hvem jeg er.....	7
Behov 2: Virke der jeg er (kanaluavhengig).....	7
Behov 3: Virke hele livet.....	8
3.3 Ulike brukergruppers behov	8
4 Menneskelige faktorer ved bruk av PID	10
4.1 Hvordan husker og gjengir vi tall og koder	10
4.2 Bekymringer knyttet til PID.....	11
Bekymring 1: Sikkerhet	11
Bekymring 2: Sammenkobling av opplysninger	11
Bekymring 3: Misbruk av PID til å undertrykke.....	12
Bekymring 4: Jeg reduseres til et tall.....	12
4.3 Brukernes egenopplevelse (spørreundersøkelse)	12
5 Brukerkrav	15
5.1 Krav til en PID.....	15
Lett å huske riktig	15
Lett å bruke riktig	15
Lett å gjengi riktig	15
Lett å oppfatte.....	15
5.2 Brukerhistorier	15
Hvordan fødselsnummer brukes på sykehus	16
I kommuneneadministrasjon.....	16
6 Referanser / kilder.....	17

1 Innledning

Dette dokumentet gir underlaget for vurdering om brukskvalitet for delprosjektet Utredning PID (Personindikator), som er en del av programmet for å fornye Folkeregisteret.

SITS Brukskvalitet er først og fremst bedt om å belyse aspekter rundt brukeropplevelse av nye Personindikator (PID) for enkeltindivider, her omtalt som *sluttbrukere*.

Kap. 2-4 utdyper bakgrunnen for kravene til en ny PID, som er i kap. 5.

Hva er Personindikator (PID)?

I dag tjener den 11-sifrede fødselsnummer som PID i det norske samfunnet: Et unikt nummer for hvert individ.

Fødselsnummeret består av 2 informasjonsenheter. Den første delen av fødselsnummeret er individets fødselsdato angitt med 6 siffer, mens den andre delen oppfattes som 5 tilsynelatende tilfeldige tall. Temporære PID (D-nummer) for personer som er nyankommet eller har midlertidig opphold har et format og oppbygning som ligner fødselsnummeret.

Når en persons PID er kjent, vet vi nøyaktig *hvilket* unik individ det gjelder.

PID alene er ikke bevis for at vedkommende er den angitte personen. Å bevise at en *er* den angitte personen, krever *autentisering*.

2 Hvem er sluttbrukere?

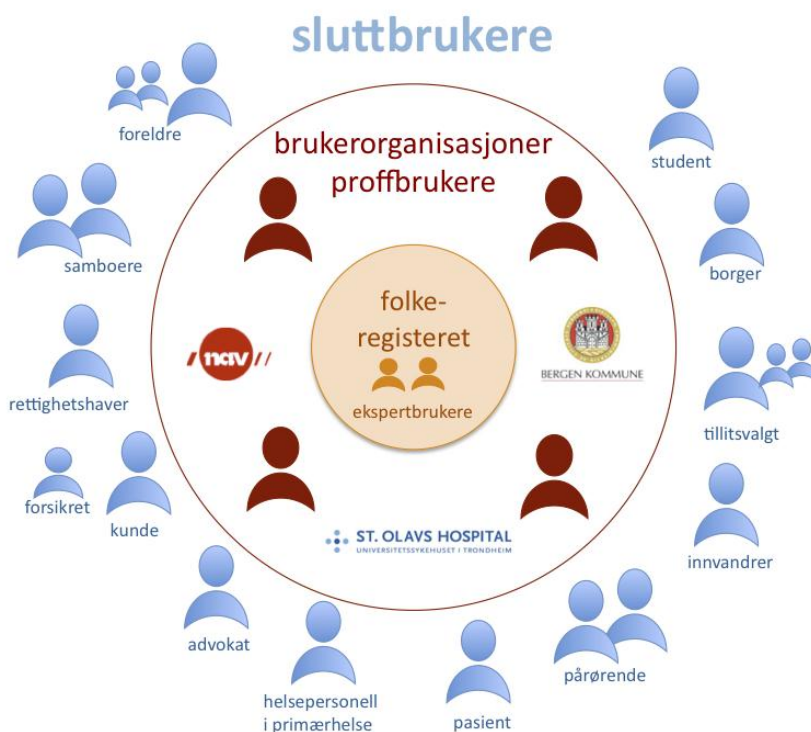
Personindikatoren (PID) benyttes av både offentlige og private organisasjoner. Disse organisasjoner er ofte omtalt som "brukere" av programmet Fornyelse av Folkeregisteret. Slike 'brukere' har mange løsninger og rutiner som benytter PID og følgelig kommer deres ansatte hyppig i kontakt med PID

Blant offentlige 'brukere' er organisasjoner som Skatteetaten, NAV, Lånekassen, kommunene og alle nivåer av helsevesenet.

Blant private organisasjoner er for eksempel banker og forsikringsselskap.

Sluttbrukerne av PID er alltid individer, først og fremst i ulike roller som proffe brukere eller som privatpersoner. Sluttbrukere av PID kan for eksempel være kunder, skattytere, velgere, foreldre, saksbehandlere, klienter, pasienter, eller borgere.

Sluttbrukerne har rettigheter, plikter eller andre interesser i forhold til de offentlige og private organisasjoner som benytter PID:



Ulike sluttbrukergrupper

En del kjente faktorer gir grunnlag for å peke ut ulike grupper blant sluttbrukere av PID:

- I hvilken grad sluttbruker har behov for å forholde seg til organisasjoner som benytter PID
- Om sluttbrukeren forholder seg til brukerorganisasjoner kun på sin egen vegne, eller også på vegne av andre: for eksempel barn, fosterbarn; syke eller demente partner eller foreldre; familiemedlemmer i utlandet
- Hvilke del av registerpopulasjonen sluttbrukeren tilhører. For eksempel: tildelt PID ved fødsel; løsere tilknytning til Norge med sikker identifikasjon (tidligere D-nummer); kjent individ med usikker identifikasjon; individ med særlig behov for beskyttelse – har hemmelig adresse, eller har måttet skifte identitet

Intervjuer av mange ulike privatpersoner vil trolig avdekke ting som ikke alt er åpenbart. I denne utredningen analysere vi kun ut fra det som er allerede åpenbart, uten intervjuer.

Minimums brukerne	Sluttbrukere med minimal behov for kontakt med brukerorganisasjonene i denne fasen av sitt liv, og uten ansvar for andre. Fikk PID med fødsel, har ikke selv bruk for det spesielt ofte i denne fasen av livet; bruker det mest i forhold til bank / forsikring.
De tunge brukerne	Sluttbrukere med utstrakt kontakt med en eller flere offentlige brukerorganisasjoner, spesielt NAV eller Helsevesenet, og ansvar for å representerer andre, hjelpetrengende familiemedlemmer overfor samme organisasjoner. (Tunge brukere av privat organisasjoner har ofte kundennummer eller kontonummer)
De fremmede brukerne	Sluttbrukere som tilhører en del av registerpopulasjonen der registrering ikke er automatisk og selvsagt. Kanskje vil de gjerne registreres, kanskje vil de helst ikke registreres. Kanskje er deres identitet ikke dokumentert på en måte som er anerkjent i Norge.
De hemmelige brukerne	Brukere med særlig behov for beskyttelse, som har hemmelig adresse, hemmelig identitet, eller ny identitet.

Ved analyse av behov og krav, er det viktig å vurdere alle identifiserte grupper, samtidig som vi husker at det kan fremdeles være viktige uidentifiserte sluttbrukergrupper.

3 Brukerbehov

Sluttbrukere av PID har forskjellige behov avhengig av om de er tilhører organisasjoner eller opptrer som individer.

3.1 Organisasjoners behov

Behovene til organisasjoner er detaljert beskrevet i andre rapporter i delprosjektet Utredning PID. Der de har lov til det innenfor rammen av Personopplysningsloven, krever organisasjoner at vi identifiserer oss entydig slik at de:

- Unngår å registrere oss som en ny bruker hver gang de har kontakt med oss eller vi kontakter dem
- Kan rasjonelt benytte den informasjonen de har om oss (f.eks. bankopplysninger ifbm skatt) ,
- Kan videreformidle informasjon om oss (f.eks pasientopplysninger fra fastlege til sykehus)
- Kan skille oss fra andre individer fordi de har likelydende navn, eller har hatt samme adresse, eller f. eks. eid samme bil.

3.2 Individets behov

Behov 1: Si hvem jeg er

Jeg har behov for å kunne enkelt og entydig angi hvilke person jeg er, når jeg er i kontakt med en offentlig eller privat organisasjon som har anledning til be meg om det.

Personnavn er menneskets uformell men godt innarbeidet måte å møte dette behovet på. Men personnavn er ikke unike. Det er tenkelig å endre norsk lov slik at det kreves at alle får et unikt navn ved fødselen eller ved ankomst til Norge. Dette vil imidlertid være en stor inngripen i den personlige friheten, og kunne føre til mye prøving og feiling i navnevalgssituasjoner.

Det er også tenkelig at brukerorganisasjonene kan benytte en kombinasjon av opplysninger slik som navn, adresse og telefonnummer for å skille individer. Dette er imidlertid ikke like sikker, siden noen av opplysningene kan endres ofte; og det er større muligheter for duplisering – at organisasjonen tror jeg er to ulike personer, fordi jeg har hatt to ulike adresser, eller endret navn da jeg ble skilt.

Behov 2: Virke der jeg er (kanalavhengig)

Jeg må kunne enkelt og entydig angi hvem jeg er i ulike kanaler: over disken, over telefon/video, på papir eller ulike elektroniske kanaler. Jeg er ikke alltid i en situasjon der jeg har kan velge kanal.

Jeg kan også havne i en situasjon der jeg ikke kan taste/lese inn min identifikator i samme takt som systemet forventer det. Dette kan føre til at jeg repeterer eller utelater deler av min PID.

Selv om vi kan forvente store endringer, vil det i mange år være behov for å kunne identifisere seg entydig i ulike kanaler. Datasystemer eller

elektroniske duppeditter kan være utslått eller utilgjengelige for den ene parten. Det må fremdeles gå an å identifisere individer.

Det er disse behovene – kanaluavhengighet, og løsning som kan fungere selv om datasystemer svikter – som gjør at PID kan ikke bare erstattes av biometri, selv om det kan være en fordel at PID kobles til biometrisk data for autentisering og for identifisering når brukere kan ikke selv samarbeide; se neste avsnitt.

Behov 3: Virke hele livet

Jeg må kunne klare å angi entydig hvem jeg gjennom mest mulig av livet mitt, uansett forbigående eller permanente funksjonshemninger, endringer i livssituasjon eller personlige karakteristikk, uansett stressede eller katastrofale situasjoner.

Spesielt når individer er syk, skadet eller i sjokk er det ofte et presserende behov å vite nøyaktig hvem de er, både for å få tilgang til korrekt medisinsk data, og for å kunne informere pårørende.

Hvis jeg er selv ute av stand til å angi hvem jeg er, fordi jeg har mistet bevisstheten eller fordi jeg er død, kan jeg fremdeles ha egeninteresse av at jeg kan identifiseres, av hensyn til mine medisinske behov (medisin allergier, kroniske tilstander, donorønsker) eller av hensyn til mine nærmeste.

Det er ikke sikkert at dette kan løses av PID direkte, men det kan ha betydning for hvilke informasjon Folkeregisteret oppbevarer om hver av oss.

3.3 Ulike brukergruppers behov

Ref kapittel 2 for definisjoner av disse sluttbrukergrupper.

Minimums brukerne	Ingen spesielle behov utover det som er nevnt for alle.
De tunge brukerne	De tunge brukerne kan ha behov for at Folkeregisteret kan ta vare på og påvise rettighetsgivende relasjoner. Disse omsorgspersoner har ofte behov for å kommunisere, handle, bestille, søke på vegne av barn, foreldre, ektefelle, partner eller samboer som er avhengig av dem. For å unngå slåsskamp og dokumentering av forhold, samtykke, fullmakt overfor hver aktør, er det naturlig at Folkeregisteret “vet sammenheng mellom PID” (Må sannsynligvis sees i sammenheng med ny vergemålslovgivning).
De fremmede brukerne	Det mest åpenbart behov til disse brukere er at Folkeregisteret er i stand til å gi PID av ulike typer til individer med ulik tilknytning til riket. Dette bidrar til å redusere antall situasjoner der utlendinger kan ikke benytte de begrensede rettigheter de måtte ha, fordi brukerorganisasjonene klarer ikke å forholde seg til individer uten kundenummer.”

	I dag har Helsevesenet måttet opprette et eget hjelpenummersystemet for å kunne håndtere individer med rett til øyeblikkelig helsehjelp, men uten fødselsnummer.
De hemmelige brukerne	Det er ønskelig for denne gruppen at det ikke er synlig for brukerorganisasjoner hvem som har hemmelig adresse eller hemmelig (eller “ny, kunstig”) identitet. Det kan kanskje bli lettere i fremtiden, da sluttbrukere med hemmelig fysisk adresse kan velge å bruke digitale kommunikasjonskanaler.

4 Menneskelige faktorer ved bruk av PID

4.1 Hvordan husker og gjengir vi tall og koder

Hukommelse er ikke avhengig av vår hensikt å huske hendelser eller informasjon. Å hente frem (huske) kan til og med skje mot vår egentlig hensikt, f.eks. du kan bare huske ansiktet og ikke navnet på en person som du selv vet at du burde kunne navnet på, sier Foster (2009) i sin bok 'Memory – a very short introduction'.

Forskere har undersøkt menneskers evne til å huske lange serier med tall. Denne forskning omfattet lengden på tallrekkefølger, evnen til å huske disse og evnen til å gjengi dem riktig. Forskningen tok utgangspunkt i situasjoner der brukere blir avkrevd lange telefonnr., bankkontonr eller personnr. I tillegg til å huske hele tallrekkefølgen må vi passe på å taste dem inn helt riktig, i riktig rekkefølge og enkelte ganger innen en tidsramme (time-out)

Cowan og Zhen (2006) har sett på hvordan vi utvikler strategier for å huske lange tall- og bokstavrekker og konkluderer med at vi tyr til meningsbærende gruppering (engelsk: chunking). Selv om slik gruppering kan gjøre det lettere å huske kombinasjonen, kommer en slik husketaktikk ofte i konflikt med systemer – som skal ha hele identifikatoren inn på et bestemt format, f.eks. uten mellomrom. Dess lengre tallrekkefølgen er, dess vanskeligere er det å huske alle tallene eller tegnene

Forskerne Cowan og Zhen (2006) sier:

It can do so through a process Miller termed chunking, the combination of multiple items to form a larger, meaningful item. To offer a simple example, it is much easier to recall the letter sequence i-b-m-c-b-s-r-c-a if one recognizes within it acronyms for three American corporations in succession: IBM, CBS, and RCA. This reduces the number of chunks in the list from nine to three.

En av de fremste norske forskerne innen fagfeltet, Knut Nordby, skrev (2001):

People are especially bad at dealing with long strings of digits. Miller's classical sevenplus-minus-two paradigm, has been much misquoted in support of introducing longertelephone numbers, viz. that most people should have no problems remembering up to nine digits (Miller, 1956). Telephone numbers, like all other numbers we have to enter (e.g. PIN-codes, security codes, passwords, etc.) must be entered absolutely correctly – every time, and the precision required by today's ICT far surpasses the criteria used by Miller and most other memory investigators.

My own research in short-term memory for numbers (using the method of immediate ordered recall for visually and auditorily presented numbers) has shown that when the number of digits in a number goes up, percent correctly recalled whole numbers falls fast (see Fig. 1): Four digits 98 % correct, six digits 87 % correct; eight digits 52 % correct; ten digits 19 % correct, and twelve digits 2 % correct (Nordby, Raanaas & Magnussen, 2001). The recalled digits must be in the right order and the whole number is incorrect if just one digit in the number is incorrectly

recalled. The reason that most people actually do manage to dial long telephone numbers correctly is that they adopt various compensatory strategies.

Aldring medfører en forverring av vår hukommelse – vi glemmer. Dette ser man i studier der personer med ellers like og sammenlignbare kognitive egenskaper, men av forskjellig alder blir bedt om å gjengi ord og bilder etter en tid. I sin studie av dette skriver Munro Cullum, Butters, Tröster, Salmon (1990) :

The deleterious effects of aging on various cognitive abilities are widely recognized, yet little is known regarding what constitutes "normal" memory test performance in individuals over the age of 74. In our study, forgetting rates for verbal and nonverbal material from the Wechsler Memory Scale-Revised (WMS-R) were examined in groups of older healthy individuals, age 50-70 and 75-95. Despite equivalent scores on measures of global cognitive status and attention/ concentration, the older group demonstrated significantly more rapid forgetting rates on the Visual Reproduction, Verbal Paired Associates, and Visual Paired Associates subtests of the WMS-R. Although patients with Alzheimer's disease also evidence very rapid forgetting on some subtests, the severity and pattern of losses appears useful in differentiating "abnormal" forgetting from that exhibited by normal elderly subjects. Preliminary normative data for normal elderly subjects on the WMS-R are presented, and the need for appropriate norms for elderly individuals is discussed.

4.2 Bekymringer knyttet til PID

Enkelte personer og miljøer er bekymret over konsekvenser av en entydig, sammenkoblet PID. De mener at det vil føre til en frarøvelse av individuell frihet og overføre makt fra enkeltmennesket til de styrende i samfunnet.

Bekymring 1: Sikkerhet

Spesielt etter terrorangrepet den 11. September 2001 ble spørsmålet om sikkerhetsaspekter ved PID aktualisert i mange land, ikke minst USA. Så å si alle de involverte terrorister var både registrert og kjent for politiet og andre myndigheter. Men den løse koblingen mellom ulike myndigheter og forskjellige registre gjorde til at opplysninger om terroristene ble ikke koblet sammen. Det råder likevel tvil blant sikkerhetseksperter om det ville ha hatt noen særskilt avgrensende virkning om bruken av entydig person hadde vært bedre, skriver Tova Andrea Wang (2002) i hennes kronikk 'The Debate Over a National Identification Card'

Bekymring 2: Sammenkobling av opplysninger

Ved at offentlige og private opplysninger kan kobles sammen kan uvedkommende (offentlige) oppdage trender og egenskaper som gir et feilaktig bilde eller avslører ting jeg ønsker å holde skjult

Helsemyndigheter og andre i forskjellige land har begynt å koble sammen helsedata for å lage helseprofiler om enkeltindivider i følge FIDIS (2009). Disse brukes av offentlige og private tjenesteleverandører til å målrette medisinsk markedsføring eller til å evaluere helseforsikring, noe som enkelte ganger har ført til at personer blir nektet forsikring.

Bekymring 3: Misbruk av PID til å undertrykke

Noen er redd for at en PID vil kunne misbrukes dersom samfunnet blir totalitært slik tilfelle var i Sovjetunionen. Myndighetenes 'gode hensikter' ble brukt som argument for så slå ned på meningsavvikere. Myndighetene kunne gjøre dette fordi de hadde samlet kunnskap om enkeltpersoner knyttet til en identifikator.

Bekymring 4: Jeg reduseres til et tall

Enkelte mennesker synes det er ubehaglig å bli redusert til en slik entydig representasjon som fratar dem alle deres personlige egenskaper som navn, kjønn, tilhørighet, m.m.

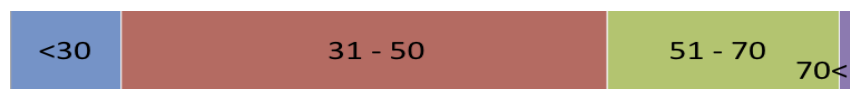
Slike entydige indentifikatorer kan få myndigheter til å bli enda mer byråkratiske og dermed fokusere på sine systemer i stedetfor menneskene i samfunnet de skal tjene

4.3 Brukernes egenopplevelse (spørreundersøkelse)

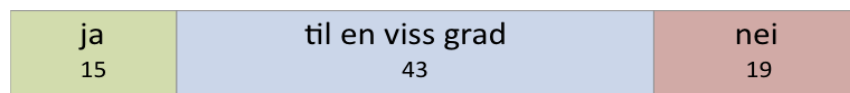
Laura Arlov utførte en enkel online spørreundersøkelse for å kartlegge egenopplevelse med å huske ting. Hensikten var å få et inntrykk av hvordan brukere forholder seg til tall og koder som de må huske.

Spørreundersøkelsen ble annonsert på Twitter / og via e-post og 77 personer valgte å delta. Denne måten å rekruttere deltakere til en spørreundersøkelse gir tilstrekkelig mengde deltakere til å betrakte svarene som valide, men de er ikke representativ for befolkningen som helhet; deltakerne må antas å være aktive nettbrukere. Allikevel kan svarene gi et inntrykk av hvordan brukere håndterer tall og andre koder slik som fødselsnummer.

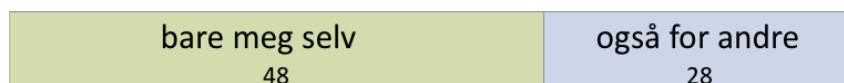
Alderen til de som svarte:



Har du lett for å huske ting som tall, passord og bilnummer?



Må du huske koder for andre enn deg selv?



Hvor mange ulike koder må du huske?

En person som ga det vanligste svaret på alle delspørsmål har minst **18 - 28** ulike passord og koder å huske på:

Vanligst svar	Passord eller kode	Hyppighet	Andre vanlige svar
1 eller 2	fødselsnummer	76%	(18 brukere må huske 3 eller flere)
1 eller 2	bilnummer	60%	(6 må huske 3 eller flere)
1 eller 2	bankkontonr	64%	(17 må huske 3 eller flere)
1 eller 2	Bank pin eller pw	58%	(31 må huske 3 eller flere)
3 - 5	Betalingskort pin	49%	(31 må huske færre enn 3) (6 må huske flere enn 5)
1 eller 2	Dørkoder på jobb	77%	(7 må huske 3 eller flere)
3 – 5	Data pw på jobb	46%	(22 må huske 1 – 2) (12 må huske flere enn 5)
6 eller mer	Nettstedspassord	65%	
1 eller 2	Privat PC passord	65%	(21 har 3 eller flere passord på privat data)

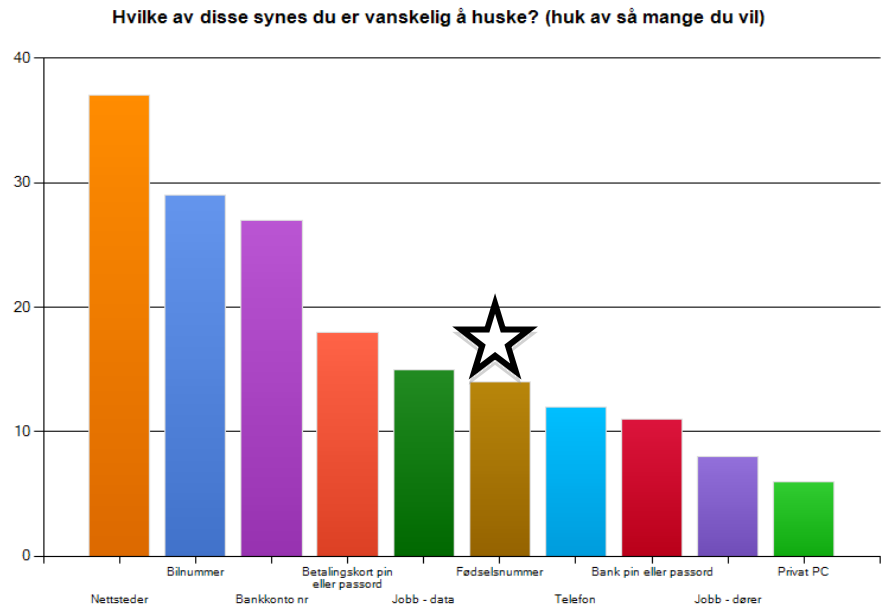
Opplever du at du skriver eller oppgir koden feil?

ja	en sjelden gang	ofte
2	56	19

25% svarer at de ofte gjør feil med slike koder, bilnummer og passord.

Hvilke koder er vanskelig å huske?

Dagens fødselsnummer kommer rimelig bra ut her, og det må være mål at en fremtidig PID scorer bedre, ikke verre enn dette:



Selv om brukere stort sett velger passord på ulike nettsteder selv, er det dette som er vanskeligst å huske.

Faktorer som karakteriserer det som er mindre vellykket enn dagens fødselsnummer er:

- Lengere enn 4 tegn og “informasjonsløse” (bilnummer, bankkontonr)
- Byttes hyppig
- Du har mange av den typen å huske
- De er sjelden i bruk

5 Brukerkrav

Brukerbehovene og hensyntagen til noen av de menneskelige faktorer beskrevet tidligere fører til brukerkrav til PID. I tillegg er noen brukerhistorier skissert for å eksemplifisere kravene.

5.1 Krav til en PID

Lett å huske riktig

Én PID. Hukommelsen er belastet nok, uten at jeg skal behøve å sjonglere med ulike ID nummer i ulike situasjoner.

En PID som er lett for meg å lære og lett å huske.

Lett å bruke riktig

En PID som jeg kan, så langt som mulig, være trygg på at jeg har oppgitt riktig, uansett om jeg oppgir den muntlig i telefon, skriver den ned for hånd eller med tastatur, eller viser frem et kort / brikke / fingeravtrykk el. lignende.

En PID som det er lov til å benytte uten autentisering i de situasjoner der dette er nødvendig.

Lett å gjengi riktig

En PID skal kunne deles inn i mindre grupper (engelsk: chunking) som gjør det lettere for meg å huske og gjengi. Det må gjerne være bestemte mønstre.

Jeg må kunne si, skrive eller taste PIDen uten besværlighet. Den bør ikke være så lang at jeg gjør feil når jeg gjengir den.

Lett å oppfatte

PIDen skal være distinkt, dvs. jeg må kjenne den igjen som min egen. Den bør også være distinkt slik at alle kjenner det igjen som en PID (og ikke blander det sammen med telefonnr, bankkontonr, osv.)

PID skal være enkelt for en mottager å oppfatte. Det betyr at en PID bør helst ikke inneholde tegn som fonetisk likeartet, f.eks bokstavene F og S, eller M og N.

En PID skal bare inneholde bokstaver og tall – ikke andre tegn som <,*,\$, m.fl.

5.2 Brukerhistorier

Noen eksempler på hvordan dagens PID (fødselsnummeret) benyttes i utvalgte sammenhenger :

Hvordan fødselsnummer brukes på sykehus

- Når jeg ringer legevakten, oppgir jeg nummeret. Sykepleieren henter frem opplysninger om meg. (Jeg er stresset. Hun ser at fødselsår og kjønn stemmer)
- Legevaktlegen skriver fødselsnummeret på epikrise og resept, og de brukes som nøkkel for reseptregisteret og for kommunikasjon til sykehuset (evt. fastlege)
- Når Per ankommer sykehuset, det første de gjør er å koble et bånd på armen hans med fødselsnummer. Hvorfor ikke bare navn? Hva hvis to Per Hansen ligger der og en skal opereres, men ikke den andre?
- Når de kommer for å hente ham til en prosedyre...Per? Når er du født?
- Årstall og kjønn i nummeret gir en stadig kontroll på at personen "matcher" nummeret.

I kommuneneadministrasjon

- Sekretær for omsorgsavsnitt i en nordlig kommune:" Vi bruker fødselsnummer for å identifisere de ulike brukerne, ja."
- Det er slik jeg vet om jeg kan putte to brukere inn på dobbeltrom for avlastning. Noen ganger kan du ikke se kjønnnet fra navnet, og det eneste stedet det står er i fødselsnummeret.

6 Referanser / kilder

Baddeley, A; Eysenck, M; Anderson, M C (2009), **Memory**

Nelson Cowan and Zhijian Chen (2006). **How Chunks Form in Long-term Memory and Affect Short-term Memory Limits**

FIDIS research program (2009). **Use and Abuse of Biometric Data and Social Networks**. <http://www.fidis.net/resources/use-cases-scenarios/use-and-abuse-of-biometric-data-and-social-networks/>

Jonathan K. Foster (2009). **Memory – a very short introduction**.

Miller, G. A. (1956). **The magical number seven, plus or minus two: Some limits on our capacity for processing information**, *Psychological Review*, 63, 81–97.

Munro Cullum C, Butters N, Tröster AI, Salmon DP (1990). **Normal aging and forgetting rates on the Wechsler Memory Scale-Revised**. University of California, San Diego, USA.

Nobuhiko Sugiura (2011). **What Needs to Be Considered When Introducing the Common National ID Number System**
<http://www.yomiuri.co.jp/adv/chuo/dy/opinion/20110207.htm>

Knut Nordby (2001). **The Expanding Telephone Number: Users' Needs for a Common Address Format in Future Converging Networks**

Tova Andrea Wang (2002). **The Debate Over a National Identification Card** http://tcf.org/publications/pdfs/pb284/National_ID_Card.pdf

Pros and Cons of a National ID Card
<http://www.apatheticvoter.com/NationalIdentityCard.htm>

Vår dato	Vår referanse
22.3.2012	2010/11979-2
Deres dato	Deres referanse
17.2.2012	
Saksbehandler:	
Jon Ølnes	

Skattedirektoratet
Postboks 9200 Grønland
0134 Oslo

Att.: Vigdis Olsen

Innspill til arbeid med ny personidentifikator i Folkeregisteret, internasjonale aspekter

Vi viser til brev fra Skattedirektoratet datert 17/2 2012 der Difi bes om å gjøre en vurdering av de framtidige utsikter for en europeisk samordning med hensyn til bruk av nasjonale personidentifikatorer og om det er forhold på europeisk nivå som kan påvirke hvordan ny personidentifikator i Folkeregisteret bør utformes.

Difis vurderinger presenteres i det følgende. Det understrekes at det er en god del usikkerhet knyttet til disse vurderingene.

Revisjon av EUs e-signaturdirektiv og relatert arbeid

Innen EU finnes i dag e-signaturdirektivet fra 1999 som regulerer spesifikt området elektronisk signatur. Dette direktivet er nå under revisjon. Det må tas forbehold om hva som kan skje under revisjonsarbeidet, men et sannsynlig utfall er at direktivet utvides til også å omfatte identifisering (navngivning, identifikatorer er en del av dette) og autentisering (bekreftelse på at oppgitt identitet er korrekt).

Den sannsynlige retningen er at alle land som omfattes av direktivet (det gjelder i dag også for EØS-området), får krav om å kunne identifisere sine innbyggere unikt innen gitt kontekst. Denne typen pålegg trenger for øvrig ikke å komme i direktivs form, men kan gis gjennom andre typer forordninger.

«Gitt kontekst» peker på at identifisering ikke trenger å være samme navn/identifikator i alle sammenhenger, og det vil neppe være krav om at navn skal kunne koples på tvers mellom ulike anvendelser. I Norge brukes én identifikator (fødselsnummer) for så godt som alle formål. Dette forenkler samspill mellom ulike aktører. Andre land kan ha andre systemer, for eksempel forskjellige identifikatorer for samme person i ulike sektorer (eksempelvis helse, økonomi/skatt, utdanning og så videre). En kan også tenke seg dette trukket helt ut til at samme person identifiseres ulikt mot hver enkelt (utenlandsk) offentlig tjeneste, men med samme identifikasjon hver gang per tjeneste.

Det er nødvendig at identifikasjonen er persistent, det vil si lik over tid. Dersom en utenlandsk borger har rettigheter eller plikter i Norge, må en kunne identifisere denne personen unikt og likt i samme kontekst over hele den perioden rettighetene/pliktene er aktive.

Autentisering innebærer at oppgitt identifisering (for eksempel en identifikator) «bevises» med tilstrekkelig grad av sannsynlighet. Det er i dag rammeverk både i Europa og i Norge («Rammeverk for autentisering og uavviselighet i elektronisk kommunikasjon med og i offentlig sektor») som bruker fire nivåer for autentisering. Dette er implementert i ID-porten i Norge med autentisering på nivå 3 (MinID) og nivå 4 (Buypass e-ID og Commfides e-ID).

I forbindelse med revisjon av direktivet er det sannsynlig at EU-land vil få krav om å tilrettelegge for autentisering av sine innbyggere overfor andre EU-land. I et direktiv må dette kravet være svært generelt formulert for å ta høyde for forskjellige nåværende og framtidige teknologier.

Krav knyttet til et revidert direktiv må antas å bli «symmetriske». Det vil si at land får krav om å kunne identifisere og autentisere sine innbyggere, men også at land får krav om å kunne akseptere slik utenlandsk identifikasjon/autentisering for tilgang til offentlige tjenester.

Prosjektene STORK og STORK-2

EUs storskala pilotprosjekter STORK (avsluttet ved utløpet av 2011) og STORK-2 (startet ved begynnelsen av 2012, for begge prosjekter se <http://www.eid-stork.eu>) er i dag referanseprosjektene for hvordan autentisering over landegrensene i Europa kan gjøres. STORK har to modeller. Den mest aktuelle for Norge er «PEPS-modellen» der brukere omdirigeres til sitt hjemland, autentiseres der, og så omdirigeres tilbake til den aktuelle tjenesten i et annet land. Et annet alternativ i STORK er «mellomvaremodellen» der et land stiller til rådighet programvare som gjør at tjenester i andre land kan tilrettelegges for å autentisere brukere fra dette landet.

STORKs løsning tilbyr ikke bare autentisering, men også utveksling av visse typer personopplysninger (definert av STORK, for eksempel alder, kjønn, adresse) mellom land, der denne informasjonen hentes fra en offisiell (i en eller annen betydning) kilde. I Norge ville for eksempel Folkeregisteret kunne være den sentrale kilden for slik informasjon.

Det finnes andre teknologiske løsninger, for eksempel Microsofts U-Prove, som kan danne basis for løsninger. Felles for løsninger er stor grad av brukermedvirkning og samtykke til utlevering av informasjon.

Det er uvisst om et nytt EU-direktiv vil dekke utveksling av identitetsinformasjon generelt eller begrenses til identifisering, autentisering og signering.

Struktur på identifikatorer og annen identifisering

Det kan tenkes at et revidert direktiv eller andre forordninger vil stille krav om at identifisering skal være gjennom én unik attributt, en identifikator. Men antagelig er det mer sannsynlig at det ikke vil være et slikt krav, noe som betyr at det vil være lov å identifisere en person gjennom flere attributter som til sammen gir unik identifikasjon. Status i Europa er varierende når det gjelder bruk av identifikatorer. De fleste land har former for identifikatorer, med Tyskland som et vesentlig unntak (i dette ligger antagelig også den største usikkerheten rundt hvor «sterke» krav et revidert direktiv kan stille).

STORK-prosjektets resultater tyder på at det ikke kan påregnes noen samordnet struktur for identifikatorer på tvers i Europa, verken når det gjelder syntaks (for eksempel hvilke tegn som tillates brukt) eller semantikk (om identifikatorer gir tilleggsinformasjon eller ikke).

Ved bruk av identifikatorer (eller identifisering gjennom flere attributter) må en derfor vite hvilket land som er kilde for identifikatoren. Dette kan gis eksplisitt gjennom at landkode prefikses til identifikator (eller attributter), eller land kan gå fram av kontekst for bruk.

I det følgende brukes begrepet «identifikator», men det bør tas høyde for at identifikasjon kan være i form av flere attributter som til sammen gir unik identifisering.

Identifisering av personer fra Norge mot utlandet

Det antas at Norge kommer til å fortsette nåværende praksis med én sektorovergripende personidentifikator. Dette burde tilsi at det heller ikke mot utenlandske tjenester er behov for å

innføre sektorvise identifikatorer. Samme person bør kunne identifiseres med samme identifikator uavhengig av hvilken utenlandsk tjeneste personen skal identifiseres overfor.

Et spørsmål Skattedirektoratet har reist, er om dette medfører at norsk identifikator bør defineres med et fast «NO-» prefiks som del av strukturen. Difis oppfatning er at dette ikke bør være nødvendig. I Folkeregisteret og for bruk mot norske tjenester bør en kunne bruke en rent numerisk identifikator som i dag.

Men dersom et «NO-» prefiks ikke er en del av selve identifikatoren, må prefiks legges til når identifikator skal brukes i en kontekst der det ikke er klart at personen er norsk. Dette bør ikke være altfor problematisk. For eksempel kan ID-porten ved autentisering mot utenlandske tjenester legge på «NO-» foran identifikatoren (i SAML-token utstedt av ID-porten). Tjenester som ber om identifikator gjennom et brukergrensesnitt, må også legges til rette for å spørre om land.

Vi nevner for ordens skyld også at personopplysningsloven regulerer bruk av entydige identifikatorer, jf. dens § 12: «Fødselsnummer og andre entydige identifikasjonsmidler kan bare nyttes i behandlingen når det er saklig behov for sikker identifisering og metoden er nødvendig for å oppnå slik identifisering.» Før fødselsnummer kan utleveres er det således et behov for en vurdering av om lovens vilkår er oppfylt. Siden de mest aktuelle løsningene i dag (for eksempel STORK) er basert på stor grad av brukermedvirkning og samtykke, kan Skattedirektoratet også vurdere i hvilken grad samtykke kan brukes ved utlevering av norsk identifikator mot utlandet.

Identifisering av utenlandske personer i Norge

Som nevnt over, kan det komme krav om at offentlige tjenester i Norge skal tilrettelegges for utenlandske personer som ikke har norsk identifikator. Svært mange IT-systemer og arbeidsprosesser er i dag avhengig av fødselsnummer/D-nummer som «brukernavn» eller identifikator i prosessene. Det vil bli meget kostbart å legge slike systemer/prosesser om til å kunne håndtere forskjellige typer identifikatorer fra forskjellige land.

Difi har i forbindelse med STORK-prosjektet (Norge er ikke med i prosjektet, men følger til en viss grad med på arbeidet) sett noe på mulighetene for å løse dette ved å etablere oversettelser mellom utenlandske identifikatorer og norsk personidentifikator. Dette er en løsning som kan vurderes i arbeidet med ny personidentifikator og modernisering av Folkeregisteret.

Dette kan gjøres ved at en på et passende trinn i angjeldende prosess allokterer en norsk personidentifikator (i dagens system ville dette vært et D-nummer) og kopler denne en-entydig til utenlandsk identifikator. «Passende trinn i prosessen» kan fortrinnsvis være når personen har oppnådd en eller annen rettighet eller plikt i Norge, men om nødvendig bør det være mulig å gjøre koplingen til norsk personidentifikator på et tidlig trinn i for eksempel en søknadsprosess. I et slikt tilfelle bør koplingen antagelig være midlertidig og slettes dersom personen gjennom søknadsprosessen ikke oppnår rettigheter eller plikter i Norge.

Dersom et slikt system for koplinger mellom utenlandsk og norsk identifikator skal etableres, må en antagelig gjøre en individuell vurdering for hvert enkelt land med tanke på hva en kan få til ut fra identifikatorsystemet i angjeldende land. Et særskilt problem kan være land som bruker forskjellige identifikatorer i forskjellige sektorer. Dette kan føre til allokering av flere norske personidentifikatorer til samme person, og disse kan ikke uten videre koples.

Ansvar for etablering og vedlikehold av koplinger mellom utenlandske og norske identifikatorer må vurderes. Dette kan ligge i Folkeregisteret, eller det kan etableres sektorvis for forskjellige tjenester, eller det kan etableres for eksempel i tilknytning til ID-porten.

Et system for kopling av utenlandske og norske identifikatorer bør også håndtere oversettelse «i motsatt retning», det vil si holde informasjon om utenlandske identifikatorer for personer i det norske Folkeregisteret. Som eksempel: Hvis en person bosatt i Norge har opparbeidet trygde-

rettigheter i et annet europeisk land, og i dette landet er identifisert ved en gitt identifikator, kan dette meldes inn i Norge slik at opplysninger lettere kan samordnes.

Gjennom etablering av slike koplinger mellom identifikatorer kan en utenlandsk person fortsette å identifisere seg med sin nasjonale identifikator og bruke e-ID fra sitt hjemland til autentisering.

Etter den informasjonen Difi har (primært fra STORK-prosjektet) om bruk av identifikatorer eller annen unik identifisering av personer i forskjellige land i Europa i dag, vil koplinger mellom utenlandsk og norsk identifikator ikke kunne gjøres generelt, men det vil kunne fungere mot en god del land. Forutsetningen må være at det andre landet er i stand til å gi en konsistent og persistent identifisering av personer, det vil samme identifikator ved hver autentisering og over tid for samme person. Dette oppfylles ikke av alle EU-land i dag, men dette kan komme som følge av et revidert e-signaturdirektiv.

Difi takker for henvendelsen og håper at informasjonen i dette svaret er til nytte. Vi ønsker et fortsatt tett samarbeid innen identitetsområdet der Skattedirektoratet med Folkeregisteret og Difi med ID-porten begge har meget viktige roller i det norske samfunnet.

Med vennlig hilsen
for Difi

Tone Bringedal
Avdelingsdirektør

Jon Ølnes
Seniorrådgiver