



Lenking og kobling i Historisk befolkningsregister

Record Linkage and Family-Relations in the Norwegian Historical Population Register

Lars Holden

Administrerende direktør, Norsk Regnesentral

holden@nr.no

Svetlana Boudko

Seniorforsker, Norsk Regnesentral

svetlana@nr.no

Gunnar Thorvaldsen

Professor emeritus og prosjektleder, UiT – Norges arktiske universitet og Historisk befolkningsregister

gunnar.thorvaldsen@uit.no

Sammendrag

Historisk befolkningsregister, HBR, er et register over den norske befolkningen fra 1801 frem til Det sentrale folke-registeret fra 1964. Det lages ved å lenke sammen personforekomster av samme person i folketellinger og kirkebøker og koble sammen nære familiemedlemmer.

Vi viser hvordan algoritmer brukes til å finne lenker og koblinger. Vi har funnet familierelasjoner i alle de nominative folketellingene, også der familiestilling ikke er beskrevet i folketellingen. Alle algoritmene baserer seg på likhet i navn, fødselsår, fødested, yrke, bosted og tilsvarende egenskaper hos familiemedlemmer. I tillegg sjekker vi at alle hendelsene gir et naturlig livsløp. Algoritmene må tilpasses de ulike kildene for å holde feilraten under en teoretisk målsetning på 1 % – noen fasit fins imidlertid ikke. Vi må for eksempel ha strengere kriterier for likhet i store kommuner som Oslo, enn i mindre kommuner og strengere i 1865-folketellingen enn i den mer nøyaktige 1891-folketellingen. Vi har funnet samme person i etterfølgende folketellinger på nasjonalt nivå ved å bruke sjeldne etternavn. Vi har også søkt etter lignende personforekomster i samme kommune og fylke i Digitalarkivet for alle personer i en av de nasjonale folketellingene.

Det er laget maskinelle lenker ved Registeringssentral for historiske data, UiT og Norsk Regnesentral. Ved nettstedet histreg.no er det også mulig å lage manuelle lenker. Vi viser utviklingen av bidragsyttere, lenker og koblinger i databasen i den fireårige perioden til registeret.

Nøkkelord

befolkningsregister, algoritmebasert lenking, folketellinger, kirkebøker

Abstract

A population register enables us to follow persons and ancestries over time. The Norwegian Historical Population Register follows the Norwegian population from 1801 until the Central Population Register takes over from 1964. It is created by linking together records about the same person in censuses and church books, making family relations explicit.

We explain how to use computer algorithms in order to find links and family relations. We have found family relations in all the nominative censuses, also in the censuses where the family position is not registered. The record link-

age algorithms are based on similarities in names, birth year, birthplace, occupation, address and relations with family members. In addition, we also check that the events make up a probable life course. The algorithms must be adapted to the characteristics of each source in order to keep the error rates low. For example, it is necessary to require stronger similarities for large municipalities such as Oslo, than in smaller municipalities.

Computer algorithms for record linkage and the encoding of family relations have been created both at the UiT – Arctic University of Norway and the Norwegian Computing Center. At the internet site histreg.no, developed by the Norwegian Computing Center and maintained by the National Archives, genealogists may see the current links and family relations and manually create new ones or change the existing ones.

Keywords

Population register, record linkage by use of algorithms, censuses, church books

Innledning

Historisk befolkningsregister, HBR, er et register over befolkningen i Norge gjennom de to siste århundrene. Registeret er under oppbygging. Det bygger hovedsakelig på de transkriberte folketellingene og kirkebøkene fra Arkivverket og Registreringsentral for historiske data (RHD) ved UiT – Norges arktiske universitet. Ambisjonen er en mest mulig komplett oversikt over den norske befolkningen fra og med folketellingen i 1801. Befolkningsregisteret har en åpen del som bare har åpen informasjon om avdøde personer, og en lukket del, som bygger på kilder som ikke er åpne. Den lukkede delen inneholder nålevende personer og er knyttet opp til Det sentrale folkeregisteret fra 1964.

De første ideene til et historisk befolkningsregister kom fra Lars Holden overfor Arkivverket i 2004 og Kåre Bævres foredrag i Demografisk forening i 2006. Prosjektet har fått finansiering fra Infrastrukturprogrammet i Norges forskningsråd fra 2010 med Gunnar Thorvaldsen som prosjektleder. Det har også vært noen andre mindre finansieringskilder. Det er laget maskinelle lenker ved RHD og Norsk Regnesentral (NR). Nettstedet histreg.no, utviklet av NR, startet opp i desember 2015, og senere i artikkelen viser vi utviklingen i antall bidragsytere og lenker fra oppstarten. Nettstedet rhd.uit.no viser lenker laget ved RHD. Registeret blir i prinsippet aldri helt ferdig, men vi kan ha en ambisjon om at en klar majoritet av personene har minst en lenke fra kilde til kilde før inneværende prosjektperiode ender i 2023. Ideelt sett ønsker vi lenker til alle forekomster av en person i Digitalarkivet og viktige referanser for personen slik som i Norsk biografisk leksikon. I praksis må vi fokusere på de viktigste kildene som folketellinger og kirkebøker og uten å legge vekt på faddere og forlovere. Når alle kirkebøker fra 1800 til 1960 blir transkribert og de fleste blir tilgjengelig i Digitalarkivet i løpet av 2020, medfører det et mye bedre kildegrunnlag for registeret.

Motivasjonen for å bygge HBR er å framskaffe en sentral nasjonal infrastruktur for forskning innen historie, samfunnsvitenskap, medisin og en rekke andre fagområder. HBR har også en viktig kulturkomponent med å følge slekter og bosetning i over 200 år. HBR er dokumentert i en rekke vitenskapelige artikler.¹ I denne artikkelen vil vi presentere metoder der datamaskiner lager *lenker* mellom forekomster av samme person i ulike kilder, og lager *koblinger*, dvs. registrerer relasjoner mellom familiemedlemmer. Dette gjøres med automatiske metoder både av RHD og NR. Vi vil også beskrive hvordan NR har tilrettelagt for manuell lenking på nettstedet histreg.no.

Lenking av personforekomster i ulike kilder har tradisjonelt vært utført i sjangeren gårds- og slektshistorie for landkommuner med detaljerte studier av de lokale kirkebøkene og folketellingene supplert med andre skriftlige og muntlige kilder.² Studiene har kartlagt

1. Thorvaldsen 2011; Holden et al. 2012; Thorvaldsen et al. 2015; Thorvaldsen & Østrem 2018; Holden & Boudko 2018.

2. Fure 2004.

bosetningen på gårdene i større grad enn personer uten fast eiendom og folk som flyttet. I de fleste tilfeller er slikt arbeid utført av personer med detaljert innsikt i de lokale forholdene. Tradisjonelt har arbeidet blitt utført manuelt, men man har i økende grad tatt i bruk datamaskiner for å effektivisere og systematisere arbeidet med kildematerialet og analysene. Maskinell representasjon forenkler også formidlingen av resultatet.³ Resultatet er vanligvis formidlet i bygdebøker.⁴

Lenking i HBR

Lenking og kobling av et nasjonalt register med 9,5 mill. personer, som levde i Norge i perioden 1801–1964 og med anslagsvis 85 mill. personforekomster i de viktigste kildene, må utføres på andre måter. For det første er det nødvendig med maskinelle rutiner som oppretter flest mulig riktige lenker mellom personforekomster. Det gjelder blant annet i kilder som har fødselsdato, for dåp av barn i samme familie med få års mellomrom i samme sogn, og mellom personforekomster i etterfølgende folketellinger når det er små endringer i familien og bostedet. For det andre er det nødvendig med et samarbeid mellom aktører som bruker ulike teknikker. Sentrale aktører som bygger HBR, bruker ulike maskinelle metoder, og vi rekrutterer frivillige til manuell lenking. Det er ønskelig å rekruttere representanter for ulike grupper og minoriteter av befolkningen for at disse skal bli best mulig dekket i registeret. Foreløpig har vi ikke lyktes med å rekruttere personer eller grupper av personer som vil fullenke et område enten internt i histreg.no eller i et eksternt program. I det siste tilfellet må lenkene leses inn i histreg.no ved hjelp av de unike IDene i Digitalarkivet. Arkivverket har skapt et nummersystem som entydig kan identifisere alle innførsler i folketellinger, kirkebøker osv. All innlesning baserer seg på disse IDene i Digitalarkivet for å unngå å lage dubletter av hendelser. Dette krever imidlertid også en bevisst bruk der både klokkerbok og ministerial er transkribert, og andre steder der flere kilder registrerer samme hendelsen. Siden klokkerboken i utgangspunktet er en kopi av ministerialboka, er det naturlig å referere til ministerialboka der begge kildene er tilgjengelige. Det fins imidlertid eksempler på at klokkerboken er mer nøyaktig og utfyllende.

Denne artikkelen beskriver den maskinelle lenking ved RHD og NR samt hvordan sistnevnte har tilrettelagt for manuell lenking i histreg.no. Vi ønsker mange bidrag fra genealoger. Noen lenker kan bare lages av trente genealoger på grunnlag av deres genealogier. Vi ønsker flest mulig lenker og at hver lenke skal ha stor sannsynlighet for å være riktig. I avsnittet om representativitet drøfter vi andre muligheter som passer bedre for noen forskningsformål. Av juridiske grunner vil mindre genealoghjelp være tilgjengelig etter ca 1930, men her er kildematerialet mer presist. I mange land må genealoger betale for å søke i slektsdatabasene. Vi er glad for at i Norge bidrar i stedet brukerne med tre ulike former for egeninnsats: De transkriberer kilder (nå især 1920-folketellinga), de kontrollerer eksisterende dataversjoner, og de utnytter sine genealogier til å lenke kildeposter.

Maskinell lenking sikrer en mer enhetlig håndtering av et stort antall kilder enn man oppnår når mange frivillige utfører manuell lenking hver for seg. Maskinell lenking kan imidlertid medføre feil som neppe ville oppstå manuelt. En algoritme som fungerer godt i ett område eller periode, kan gi færre riktige lenker og flere feil i et annet område på grunn av mye flytting, egenskaper ved kildene, navnetradisjoner eller andre forhold. Noen lenkingsalgoritmer baserer seg bare på informasjon i to personforekomster uten en vurdering

3. Kjelland 2016.

4. Bygdebøker, hentet fra <https://lokalhistoriewiki.no/>

av andre personer eller kilder fra hele livsløpet. Da må dataene være relativt unike og konsekvente med litt sjeldne navn, fødselsår og bosted i to folketellinger, eller med samme navn og fødselsdato. Heldigvis kan man ofte supplere med data om andre familiemedlemmer og for eksempel bygge livsløp der man finner igjen søsken med samme navn og alder i kirkebok og folketelling. En vanlig teknikk er å bruke ektepar med stabile navn som fikk flere barn i samme prestegjeld. Lenkingsteknikkene er et kompromiss mellom å lage så mange riktige lenker som mulig samtidig som man ikke introduserer feilaktige lenker. Det er derfor nødvendig å legge inn kontroller for å sikre at samlingen av alle personforekomstene som er lenket sammen, gir realistiske livsløp. Det gjelder og ved manuell lenking. For eksempel må det sjekkes at barnefedre var i live ni måneder før fødselen og at mødre var i fruktbar alder.

RHD har utviklet programvare for automatisk lenking av individposter fra ulike kilder som omhandler samme person i databasesystemet Oracle. Denne egenutviklede løsningen viser seg mer fleksibel og slagkraftig enn bruk av standard programvare for lenking. Den gir høyere lenkingsrater enn for eksempel vår samarbeidspartner Minnesota Population Center oppnår med sin standardløsning. De automatiske lenkene overføres til histreg.no hvor vi forventer at genealoger kan bringe andelen lenker over 80 %. Det er skapt ca 5 millioner lenker mellom transkriberte personposter fra kirkebøker og især folketellinger i perioden 1801 til ca 1920. Lengst har RHDs lenkingsarbeid kommet i Troms. Når vi får tilgang til de kirkebøkene som nå blir transkribert i utlandet, vil dette tallet øke vesentlig også i andre deler av landet. For å unngå skjevheter i utvalg av lenkede data er det utviklet nye lenkingsrutiner for single individer, som nå settes i produksjon. Tidligere lenker var oftest basert på par-relasjoner.⁵ Vi forventer relativt færre lenker i byene og relativt flere mellom personforekomster i nyere kilder.

Også lenkingsalgoritmene til NR er i stor grad basert på en scorefunksjon der man får poeng for likhet i fornavn, etternavn, fødselsår, bosted, fødested, yrke og å gjenkjenne familiemedlemmer. I tillegg tas det hensyn til livsløp når det gjaldt forventet alder for konfirmasjon, vielse og foreldre i dåp, konfirmasjon og vielse. Fornavn splittes opp i separate navn, og likhet i navn måles med en datarutine utviklet av Jaro og Winkler. Jaro-Winkler er et avstandsmål for likhet mellom to ord basert på antall sammenfallende bokstaver og deres rekkefølge. Et kritisk punkt er vektning av de ulike komponentene i scorefunksjonen og å bestemme terskelen for hvor mange poeng som trengs for å godkjenne lenking av personforekomster.

Noen lokale lenkingsdatabaser rapporterer om opptil 90 % lenkingsgrad, men forfatterne er ofte ikke presise på hvordan dette er definert. Vi har et potensiale for å klare det bedre enn dette fordi vi også kan lenke sammen personer som flyttet innenlands. Men vi har ikke mulighet til å legge like mye arbeid i hver lenke som lokalhistorikere kan når de fullenker et område. Dessuten vil det være lettere å oppnå høy lenkingsgrad i mindre bygder med lite migrasjon, enn i byer med mange mennesker og stor mobilitet. Flere kilder er sparsomme. Tidlige begravelser ble ført med bare fornavn og etternavn. Da har man i de fleste tilfeller ikke nok informasjon til å lenke personer individuelt, når vi krever at hver lenke har høy sannsynlighet for å være riktig. Det hjelper om det i tillegg fins fødested, fødselsår og familiemedlemmer, men i store byer er heller ikke dette alltid tilstrekkelig. Hvis kildegrunnlaget er godt, slik at man har god kontroll over hele befolkningen, kan man i større grad utelukke at en bestemt personforekomst kan være andre enn den aktuelle lenkingskandidaten. Det er mulig å beskrive at en lenke er usikker. Hvis en personforekomst, A, i en begravelse kan tilhøre to forskjellige personer i bygdas folketelling med samme navn,

5. Bouchard 1992.

B og C, kan vi ikke uten videre lenke personforekomsten A som usikker til både B og C, hvis dette ville medført en sammenlenking av B og C.

Oppbyggingen av histreg.no og manuell lenking

Histreg.no er en overbygging til Digitalarkivet, og de unike IDene for hver personforekomst er svært sentrale. Disse IDene brukes både i internettadressene i Digitalarkivet og i personsidene i histreg.no. Histreg.no har en database med lenker mellom personforekomster i ulike kilder om samme person. Lenkene er representert i et stjerneformat der en hovedpersonforekomst er lenket til hver av de andre personforekomstene. I tillegg er det lister med familierelasjoner mellom partnere, foreldre og barn og mellom søsken der man ikke kjenner foreldrene. Det er mulig å representere foster-, ste- og adoptivbarn. I histreg.no har hver person en personside, se figur 1, som viser alle personforekomstene som er lenket sammen. Her er det mulig å redigere navn, fødsels- og dødstidspunkt og sted. Det er mulig å skrive en biografi med begrunnelse av lenkingen og lage referanser til andre kilder. Det er blant

Johanna Halvorsdatter Løkken

Kvinne
Født: 07.08.1819 , Røros
Død: 06.02.1900, Røros

Søk etter lignende personer

Foreldre og søsken

Far: [Halvor Johnsen Unsgaard Løkken](#), 1781 - 1877

Mor: [Ingeborg Andersdatter Haugen](#), 1787 - 1830

Søsken: [Bereth Lucia](#), 1818 -

Søsken: [Anders Halvorsen](#), 1821 -

Søsken: [Ingeborg](#), 1828 -

Partner og barn

Partner: [Anders Larsen Holden](#), 1821 - 1870

Barn: [Halvor Andersen Holden](#), 1849 -

Barn: [Peder Holden](#), 1852 - 1935

Barn: [Johan Andersen Holden](#), 1854 -

Barn: [Johannes Olsen](#), 1864 -

Barn: [Johannes Olsen Funneshdahl](#), 1863 -

Livsløpstabell - Oversikt over kilder med personen

Nr	Dato	Kilde	PFID	Rolle	Navn	Fødselsdato	Fødested	Bosted	Fam. stil.	Sivst.	Yrke	Far/brudgom	Mor/brud	Partner
1	19.09.1819	MINI , 0434...1643 , 1817-1828	pd00000004253769	barn	Johanna	07.08.1819						Halvor Jønsen Unsgaard	Ingeborg Anders Datter Haugen	
2	31.12.1865	FOLK , 1640...1643 , 1865-1865	pf01038322001806		Johanna Halvorsdatter	1819	Røraas	Røraas: Bakkene*	hans Kone	g				
3	31.12.1875	FOLK , 1640...1643 , 1875-1875	pf01052371002003		Johanna Halvorsdatter Holden	1819	Røraas	Røros: Røraas	hm	e	Selveierske			
4	01.01.1891	FOLK , 1640...1643 , 1891-1891	pf01053088002640		Johanna Holden	1819	Røros	: Peder Holden	moder	e				

Biografi og merknader

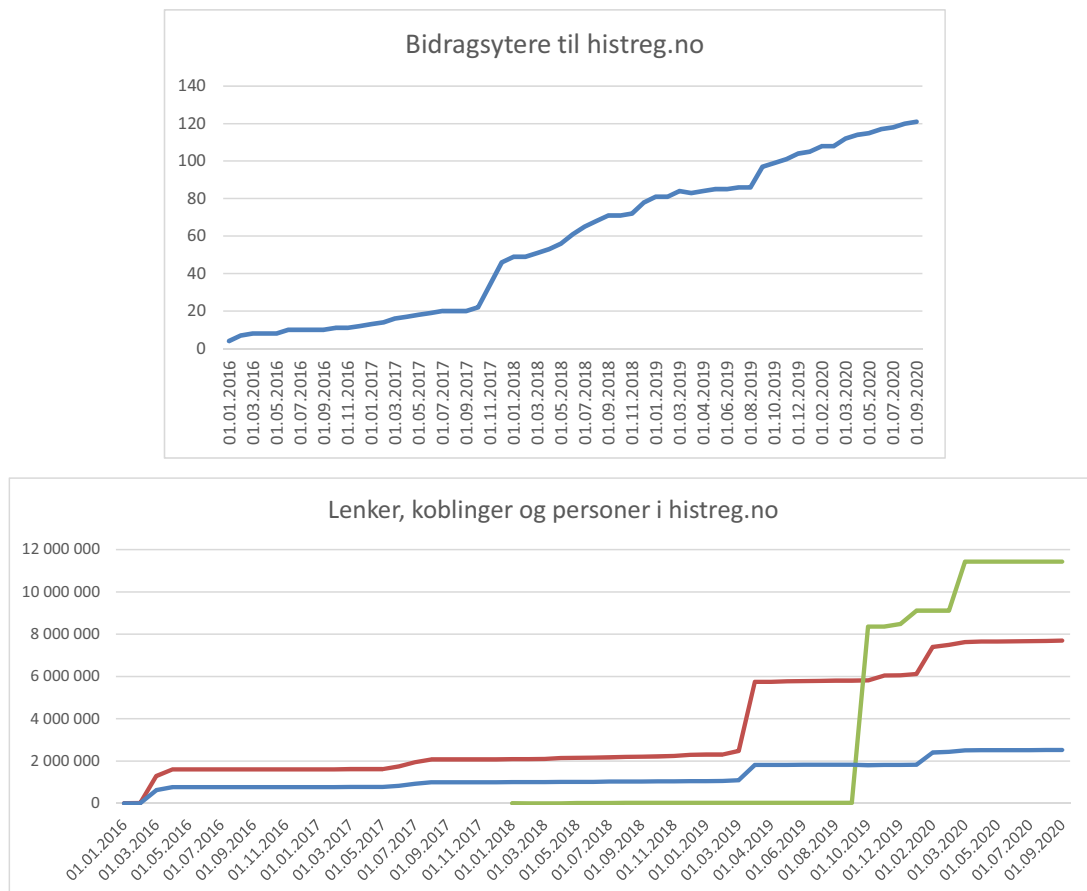
Hun og Anders giftet seg 1844. Både Anders og Johanna var vaksinert mot kopper, han i 1828, hun noen måneder gammel. Johanna hadde barna: Iver, Halvor, Peder, Johan,

Figur 1 Figuren viser personsiden til tippoldemor til en av forfatterne. Vi har funnet kilden for hennes egen dåp og i folketellingene 1865, 1875 og 1891. Johannes Olsen var fostersønn i 1865-folketellingen. Vi ser hvordan navnet endret seg i de ulike kildene. Først etter mannens død, benyttes hans etternavn. Hennes far bodde hos henne som enkemann i de to første folketellingene. I de to siste var hun selv enke, og i den siste bodde hun hos en av sønnene. All tekst i blått er lenker til andre personsider, Digitalarkivet eller Lokalhistoriewiki. Det er en kort biografi som bl.a. forteller at eldstebarnet Iver flyttet til USA i 1863, og en kort beskrivelse av hans familie.

annet et stort antall lenker til biografier i Lokalhistoriewiki som er laget automatisk basert på IDene i Digitalarkivet. Alle bidrag til histreg.no krever at man registrerer seg som bruker med fullt navn. Alle bidrag blir logget med bidragsyter og tidspunkt for bidraget.

Søk i histreg.no er et søk i alle transkriberte kilder i Digitalarkivet og ca 0,5 mill. hendelser i aviser, hovedsakelig dødsannonser fra Aftenposten. Det er mulig å sortere søket alfabetisk etter fornavn, etternavn, fødested, bosted, kilde eller kronologisk etter fødselsår og hendelsesdato. Videre er det mulig å vise navn til foreldre og partnere, hvis disse er lenket. Dette bidrar til at det er mye lettere å identifisere samme person i søket. Det er mulig å lenke direkte fra søket. Det er også mulig å be datamaskinen finne hvilke grupper av personforekomster som bør lenkes sammen. Bidragsytere kan ved et klikk akseptere, modifisere eller forkaste den foreslåtte gruppen med personforekomster som kan lenkes.

Figur 2 viser utviklingen i antall bidragsytere, lenker, familiekoblinger og personer med lenker i histreg.no. Bidragsyter er definert som en person som har registrert informasjon i histreg.no. Antall lenker er definert som antall personforekomster som er lenket til minst en annen personforekomst. Ideelt sett ønsker vi dermed at dette skal være nærmest mulig antall personforekomster i de tilgjengelige kildene, dvs. Digitalarkivet pluss registrerte hendelser fra avisene. Koblinger er registrering av familierelasjoner mellom personer som ikke eksplisitt følger direkte fra en kilde. Alle familierelasjoner som er direkte definert i en kilde, for eksempel mellom barn, mor og far i en dåp, hentes fra Digitalarkivet og regnes derfor ikke som en kobling. I en periode frem til september 2019 ble familierelasjoner i 1801- og 1910-folketellingene heller ikke regnet som koblinger. Fra denne datoen ble disse mer løst definerte familierelasjonene lest inn i histreg.no slik at de manuelt kan redigeres og regnes som koblinger.



Figur 2 Figurene viser vekst i bidragsytere og lenker (rød), koblinger (grått) og lenkede personer (blått) i histreg.no fra starten i januar 2016 og frem til 1.3.2020.

Oppslag i aviser og radio høsten 2019 medfører at en stor andel av personer, som er så interessert i slektsforskning at de går på slektsforskerdagene, kjenner til histreg.no. Det er derfor grunn til å tro at antall brukere er vesentlig høyere enn antall bidragsyttere. Erfaringen viser at innlegg i Arkivverkets debattforum er mer effektivt for å rekruttere bidragsyttere enn radioprogram der man når over 100 000 personer. Hoppene i antall lenker frem til sommeren 2019 skyldes innlesning av lenker laget ved RHD.

Representativitet

Målsetningen er som nevnt å lage flest mulig lenker mellom personforekomster av samme person i ulike kilder og koblinger mellom personforekomster av familiemedlemmer nevnt i ulike kilder der hver enkelt lenke og kobling er av god kvalitet. I noen forskningsprosjekter kunne man ønsket seg et svakere krav til kvaliteten av hver lenke for å kunne følge en større andel av befolkningen. For å illustrere: For noen typer forskning vil det være bedre å kunne lenke en Ole Olsen fra Rendalen i en emigrantprotokoll eller en begravelse til en av de døpte med dette navnet i bygda enn at denne personforekomsten er ulenket. Dette kan imidlertid skape feilaktige resultater når livsløp blir satt sammen av data som egentlig tilhører ulike personer.⁶ Det er en balanse mellom å lage mange lenker og at lenkene skal være representative der ulike forskergrupper har ulike behov og metoder.

Histreg.no brukes av slektsforskere. For dem er det viktig at hver enkelt lenke er meget sannsynlig, ikke bare at lenkene samlet er representative for befolkningen. For noen typer medisinsk forskning er dette også viktig – for eksempel kan feil i lenking av fedre forstyrre studiet av arvelighet. For forskere er det ønskelig med en presis beskrivelse av hvilke lenker som er inkludert, for eksempel alle lenker laget med samme algoritme og med kjente parametere: navn, alder, fødested osv. Det er ikke mulig å oppnå når vi er avhengig av manuelle bidrag fra mange personer. Erfaringene fra maskinell lenking er også at man arbeider skrittvis med stadige utvidelser og justeringer av programvaren basert på nye kilder og nye erfaringer med lenking.

I den maskinelle lenkingen har vi et mål om at det skal være under 1 % feil, men det er svært vanskelig å vurdere dette. Det vil være en balanse mellom tilfredsstillende lenkingsgrad og andel feil. Vi har ikke fasit, og det kan være stor variasjon i kilder og navnetradisjoner. Det som virker som en god lenkingsalgoritme noen steder, kan gi mange feil andre steder. Det som kan virke som et meget sjeldent navn, kan være et vanlig navn, som mange søskenbarn hadde. Derfor tar gode lenkingsprogram hensyn til hvor stor andel av befolkninga som har ulike navn, fødesteder osv. En av de vanskeligste avgjørelsene ved maskinell lenking, er å sette score-terskelen for når vi skal regne en lenke som tilstrekkelig sikker til å kunne inkluderes.

Vi vil få en vesentlig større lenkingsandel blant personer med utdanning og formue, bofaste, med mange barn og sjeldne navn. Det vil være spesielt vanskelig å dekke omstreifere og andre med stor mobilitet og vanskelige familieforhold. De manuelle bidragene vil i stor grad bidra til denne skjevheten, men også enhver algoritme som gir et akseptabelt antall lenker, vil gi slike skjevheter. Flere eksperter argumenterer⁷ for at man ikke skal bruke adresser eller informasjon om yrker for å redusere denne skjevheten, bare egenskaper som ikke endret seg gjennom livsløpet. Men hvis man utelater informasjon i lenkingen, vil man få færre lenker og større andel feil. Vi mener derfor at det er best å bruke all informasjonen og lage flest mulig lenker av god kvalitet. Da aksepterer vi også at det er en skjevhet i hva som lenkes. Det er flere grunner til dette valget:⁸

6. Ruggles et al. 2017.

7. Christen 2012.

8. Disse problemstillingene er diskutert blant annet i Thorvaldsen 2006; Goeken, Huynh et al. 2011; Ruggles et al. 2017.

- Enhver lenkingsstrategi vil gi skjevheter slik at man uansett ikke vil få like stor lenkingsgrad i alle deler av befolkningen.
- Kriteriene for lenkingen kan flagges. Forskere som ikke ønsker å basere sine resultater på manuelle lenker eller lenker laget vha. yrker og adresser, kan da fjerne disse fra sine datasett.
- Den beste måten for å få representative utvalg og statistikk er å korrigere ut fra frekvenser i folketellingene. Det er mulig å korrigere for bosted, familieforhold og yrke basert på frekvens i folketellingene. Dette er nærmere beskrevet i herværende artikkel om brukergrensesnitt til HBR.⁹
- For å være viktig for slektsforskere og få manuelle bidrag fra denne gruppen, er det viktig at registrerte lenker er sannsynlige. Det vil gi en skjevhet i hva som registreres, noe som kompenseres med merking av lenkenes opphav.

Familierelasjoner fra folketellinger

Den maskinelle lenkingen og koblingen som er utført av Norsk Regnesentral, er basert på programmering i statistikkverktøyet R og med tilgang til kopier av folketellingene.

Folketellingene er registrert per husholdning, men det er ikke alltid lett å tolke familieforholdene i husholdningen. Vi ønsker å identifisere familieforhold som partnere og foreldre-barn, inklusiv ste-/foster-/adoptivbarn. De fleste familierelasjonene er mellom hovedpersonen (husfar eller husmor) og deres barn. Men det er ikke uvanlig at foreldre og søsken til husfar og husmor er med i husholdningen. Vi ønsker også å identifisere søsken uten at man kjenner noen av foreldrene. Mer fjerne forhold som barnebarn og svigerfamilie, registreres ikke hvis vi ikke identifiserer de personene som knytter dem sammen. Det er forskjeller mellom folketellingene og lokalt innen en folketelling om hva som er registrert i folketellingen. I 1910-folketellingen er de aller fleste i kjernefamilien registrert med hf, hm, s og d for henholdsvis husfar, husmor, sønn og datter i familiestillingsfeltet. Da kan man identifisere de fleste i kjernefamilien med en enkel algoritme. En algoritme er en enkel oppskrift av denne typen: Hvis etterfølgende personer i en folketelling har familiestillingene «hf» etterfulgt av «hm», «s» og «d» tolkes dette som far og mor i en familie med to barn. I de eldre folketellingene er det større variasjon i hvordan disse familiestillingene er beskrevet og andre familiestillinger som husfar/husmor sine foreldre og søsken. Dette er løst ved å finne 20–70 forskjellige synonymer for disse familiestillingene f.eks. «husfar», «husbond», «huusbond», «hovedperson» etc. For noen koblinger har vi også brukt flere felt, som sivilstand og alder, i tolkningen. Flere ord har tvetydig betydning. For eksempel er ordet «moder» tolket som husmor om vedkommende og husfar er gift og omtrent på samme alder. Hvis aldersforskjellen er over 20 år, regnes det som mor til husfar, og om det ikke er husfar i familien som mor til husmor. Det er også andre utfordringer som at et av barna har familiestilling «s hf» og neste person har familiestillingen «kone». Dette paret oppfattes da som sønn og svigerdatter av husfar og husmor hvis disse er nevnt tidligere i husholdningen. I folketellingen 1865 er familiestilling og yrke skrevet i samme felt slik at man må forsøke å tolke deler av dette feltet. I noen folketellinger brukes ofte begreper som «sønn av nr 8», der det henvises til person nummer 8 i husholdningen.

Vi har brukt patronymikon i gjenkjenning av familierelasjonene for de folketellingene der vi mener det gir bedre resultat. Praksisen for bruk av patronymikon i folketellinger varierer mye mellom og innen tellingene fra slutten av 1800-tallet. Dette gjør det vanskelig å vite hvilken algoritme som gir best resultat.

9. Thorvaldsen et al. 2020.

Ideelt sett skulle man jobbe lenge med algoritmen og så bruke samme algoritme på alle folketellingene. I praksis må det arbeides iterativt med en og en folketelling. Programmet utvikles med hovedsakelig testing mot en folketelling. Programmet brukes på denne folketellingen når vi tror andel feil er tilstrekkelig lav. Koblingene leses inn i befolkningsregisteret slik at det hjelper alle som arbeider med manuell lenking på histreg.no. Deretter gjentas dette for hver av de andre folketellingene, med uttesting og forbedring av algoritmen til feilraten er tilstrekkelig lav før koblingene lages og leses inn i befolkningsregisteret. Manuell tolkning ville gi flere familierelasjoner enn dataprogrammet finner, kanskje rundt 10 % og høyest for de eldre tellingene. Vår ambisjon har vært å finne flest mulig av de lette koblingene og begrense feilene. Brukere av histreg.no kan lett slå opp i folketellingene i Digitalarkivet og kontrollere hva som er riktig. Bidragsytere kan legge til nye relasjoner og rette opp feil.

Tabell 1 gir en oversikt over de transkriberte folketellingene som er tilgjengelige i Digitalarkivet. Tabellen viser i kolonnen «famst» antall poster i tellingene og prosentandel av postene der familiestillingsfeltet eller yrkesfeltet er utfylt. Vi har valgt å ta med begge feltene fordi folketellingen 1865 har registrert familiestilling sammen med yrkesfeltet. Andre felt kan også være viktige for å finne familierelasjonene. Flere av disse kildene er kladder til de numeriske folketellingene fra 1815 til 1855. Det er tilfeldig for hvilke prestegjeld disse kildene er tilgjengelige og hvordan de ble ført, for disse nominative listene var bare prestenes kladder. Tabellen viser hvor mange ulike familiestillinger vi har lyktes med å gjenkjenne. Det viser hvor vanlig det er med andre familiemedlemmer som bor sammen med kjernefamilien. Det må imidlertid tas forbehold om hvor mye vi har klart å gjenkjenne. I 1865-folketellingen kan en familiestilling et stykke ut i husholdningen være «inderst», «losjerende» eller tilsvarende betegnelser etterfulgt av «kone» og eventuelt barn. Det var nødvendig å inkludere disse ordene som husfar for å klare å få omtrent like mange husfar som husmor. Hvis noen tall er overraskende små, kan det skyldes at vi bare har funnet en liten andel av personene med denne familiestillingen. Generelt vil det være vanskeligere å gjenkjenne mer kompliserte roller, som fosterbarn, fordi dette kan beskrives på mange ulike måter. Hvilke felt som er fylt ut, varierer mye mellom folketellingene. For eksempel har flere numeriske folketellinger ikke oppgitt fødselsår, og alder er bare oppgitt som intervaller, og 1825-folketellingen har som oftest ikke fylt ut sivilstand.

I noen av folketellingene er familiestilling mangelfullt registrert som vist i Tabell 1, kolonne 3. Da har vi valgt å bruke noen regler som i de fleste tilfeller gir riktige familieforhold: Mann etterfulgt av kvinne som begge er gift og med under 15 års aldersforskjell, antas å være gift med hverandre. Personer som følger etter hverandre i folketellingene, har samme etternavn, er ugift eller har samme kjønn og aldersforskjell på under 6 år, antas å være søsken. Personer som er under 14 år antas å bo sammen med minst en forelder. Vi antar at etterfølgende personer med samme etternavn og der den første er over 18 år eldre, er foreldre og barn. Slike kriterier gir noe feil i gruppehushold som sykehus og skoler.

RHD har kodet familie- og husholdsrelasjonene i henhold til det internasjonale systemet utviklet ved Minnesota Population Center (MPC) for enkelte av folketellingene, se nettstedet ipums.org. I tillegg har MPC laget såkalte lokasjonspekere som viser hvilke familiemedlemmer som hører sammen. Som regel vil ektefellene stå først i husholdet og ha ektefellepekere til hverandre. Barna vil som regel ha pekere som viser til faren og til moren. I mer kompliserte hushold er disse pekerne svært nyttige for å identifisere relasjoner mellom familiemedlemmene eksplisitt. De amerikanske programmene har dessverre den svakhet at de ikke tar hensyn til patronymikon.

Tabell 1 Alle tall bortsett fra det som er oppgitt i prosent og årstall, skal ganges med 1000. Kolonnene fra venstre er antall nominative poster (personinnførsler) i folketellingen tilgjengelige i Digitalarkivet, prosentandel poster med utfylt familiestilling eller yrke (Famst), antall husfedre (Hf), husmødre (Hm), barn, barn som er adoptiv-, foster eller stebarn, husfars foreldre (HfF), husmors foreldre (HmF), husfars søsken (HfS), husmors søsken (HmS) og antall koblinger laget fra folketellingen. LFOT er lokale folketellinger i 9 prestegjeld og KFOT 8 kommunale folketelling hvorav 7 byer.

År	Poster	Famst	Hf	Hm	Barn	Adoptiv	HfF	HmF	HfS	HmS	Koblinger
1769	4,5	15 %	0,6	0,2	2,0	0,1					3,0
1801	879	98 %	136	141	327	0,7	6,3	4,7	3,5	2,3	675
1815	67	35 %	9	8	20	0,2					36
1825	31	47 %	4,8	4,1	11,4	0,2					20
1835	14	58 %	2,0	2,2	6,2	0,08					11
1845	12	70 %	1,5	1,4	4,4	0,07					8,0
1855	1	46 %	0,2	0,2	0,5						0,8
1865	1 688	83 %	346	276	703	17	11,3	2,2	4,1	0,6	1 761
1875	1 830	99 %	377	306	707	24	3,5	2,7	4,4	0,9	1 828
1885	443	6 %	36	43	169	1,9	0,04	0,08	0,1	0	352
1891	1 937	94 %	354	270	790	21	2,4	2,0	5,9	0,2	1 736
1900	2 316	98 %	372	398	1 027	18	1,7	2,4	3,1	1,2	2 241
1910	2 475	98 %	411	439	1 145	18	1,4	1,9	2,8	1,2	2 462
LFOT	13	95 %	1,6	1,2	6,2	0,1					10
KFOT	274	0,4 %	59	43	85	1					225

Lenking mellom folketellinger basert på sjeldne navn

Etter å ha beskrevet rutiner for kobling av familierelasjoner, vender vi nå tilbake til noen mer spesielle teknikker for lenking av forekomster av samme person i flere ulike kilder. NR har laget en egen algoritme som lenker mellom personforekomster med sjeldne etternavn i to etterfølgende folketellinger. I programmet R er det lett å finne frekvensen av alle etternavn når hele folketellinga er lest inn i et dataobjekt. Vi identifiserer lenker mellom personforekomster ut fra følgende algoritme, litt grovt beskrevet:

1. Finn alle forekomster av hvert navn der vi inkluderer nesten like navn ut fra Jaro-Winkel med terskel som omtrentlig aksepterer høyst en bokstav forskjellig pr 6 bokstaver i navnet.
2. Hvis det totale antallet av dette navnet inklusive varianter er under 3000, forsøker vi å finne lenker.
3. Vi sammenligner alle personforekomstene i den ene folketellingen mot alle i den andre folketellingen der begge har det samme etternavnet. Vi gir en score for likhet ut fra fødselsår, fornavn splittet opp i opp til 4 navn som sammenlignes parvis, etternavn, fødested og bosted.
4. Der scoren i punkt 3 kommer over en terskel, ser vi om flere i samme husholdning kommer over terskelen. Dette gir i tilfelle ekstra score til lenken.

5. Vi lager så lenker der scoren i punkt 4 er over en terskelverdi og avstanden i scoren til nest beste match er over en terskelverdi.

Når vi sammenligner navn, fødested og bosted, er dette en sammenligning bokstav for bokstav uten å ta hensyn til alternative skrivemåter av samme navn eller geografiske avstander. Når vi ser på likhet mellom personforekomstene for andre familiemedlemmer, ser vi bort fra familiestilling. Det følger ofte av alder i forhold til den personen vi vurderer å lenke. Det gjør at vi også får med oss personer i kompliserte familieforhold som er vanskeligere å oppfatte maskinelt som losjerende hos foreldre og søsken. Når det er to gode matcher, sjekker vi om det skyldes at personforekomsten er midlertidig til stede og fraværende i de to forekomstene. Hvis det er tilfelle, regnes begge som en match. Det er viktig å ta hensyn til nest beste match. Det gjør at kravet for match blir høyere for vanlige etternavn enn for mer sjeldne etternavn. For sjeldne etternavn kan det være tilstrekkelig med samme fødselsår og et lignende fornavn for å lenke sammen personforekomstene. For vanlige navn er det ofte et krav om flere like fornavn og familiemedlemmer. Å utnytte sjeldne navn bidrar til flere lenker og mer interesse fra genealoger, men mindre representativitet.

Vi ser at vi ofte klarer å lenke mellom barn i første folketelling til voksne i neste folketelling. Vi ser at bruk av etternavn er vesentlig vanligere i senere folketellinger enn i de første. I denne lenkingsmetoden er det spesielt vanskelig å finne gode terskler ifølge punkt 5. Vi kan gjøre noen stikkprøver, men det er i de fleste tilfeller vanskelig å vite fasiten. Det kan være store forskjeller mellom ulike etternavn. Tabellen nedenfor viser antall lenker mellom folketellingene:

Tabell 2 Resultat av lenking mellom etterfølgende folketellinger basert på sjeldne navn.

Lenking mellom år	Første folketelling				Antall lenker	Lenker som var der allerede
	Antall navneformer (nesten like navn slått sammen)	Antall personer	Personer med navn med under 1000 forekomster	Personer med navn med mindre enn 20 forekomster		
1801–1865	10 324	878 978	45 455	28 238	692	< 10 %
1865–1891	16 711	1 688 306	130 268	55 364	9 588	< 10 %
1891–1900	41 763	1 936 553	647 611	139 141	82 378	< 10 %
1900–1910	52 861	2 315 654	899 518	188 583	148 689	ca 70 %

Lenking mellom kilder med fødselsdato

Noen store kilder kan vi behandle spesielt. Det gjelder lenking mellom folketellingene 1910 (2,4 mill. personer), listen med døde mellom 1951–2014 (2,5 mill. personer) som ble lagt ut i Digitalarkivet i april 2020 og folketellingen 1920 (2,6 mill. personer) som kan offentliggjøres i desember 2020. Alle disse kildene har fødselsdato og egner seg dermed godt for maskinell lenking. I prinsippet er lenkingsmetoden lik, men konsekvent ført fødselsdato gir mange flere og sikrere lenker. RHD har lenket familier ved å identifisere barn under to år som hadde fødselsdato i 1900-tellinga igjen i folketellinga fra år 1910 da alle for første gang ble ført med fødselsdato. Det er laget lenkingsalgoritmer som fokuserer på feil i fødselsdato og personer som skiftet etternavn. Stikkprøver mot dåpslistene antyder at minst 10 % i 1910-folketellingen hadde feil fødselsdag og at opptil 40 % kan ha feil i deler av datoen i lokale tellinger. Det er ikke uvanlig at samme person har ulikt etternavn i ulike kilder:

patronymikon, gårds- og stedsnavn, familienavn og mannens etternavn i tillegg til variasjoner i staving av navnet. Bruken av patronymikon fortsetter helt frem til 1920-folketellingen på landet. Generelt gjelder at navnebruken i like stor grad bestemmes av kildeføreren som av personen selv.

Lenking ved søk lokalt i Digitalarkivet

Et stort antall lenker kan lages ved å gjenfinne personforekomster i ulike kilder i hvert lokalmiljø automatisk. Det er nå ca 35 mill. personforekomster i Digitalarkivet, og vi regner med at dette stiger til ca 90 millioner personforekomster når alle kirkebøkene etter 1800 blir transkribert i løpet av få år. Det er også mange andre kilder i Digitalarkivet med informasjon om emigrasjon, skatt, skifterett, fengsel, vaksiner osv. Avhengig av hvor fullstendig informasjon disse inneholder, er de mer eller mindre kurante å lenke maskinelt. Vi har gode erfaringer med dåp og konfirmasjon familievis ut fra navn på far, mor og bosted og som sammenfaller med familien i en folketelling. Det legges stadig ut nye kilder i Digitalarkivet, og vi ønsker en metode som raskt kan etablere de lette og mest relevante lenkene når nye viktige kilder legges ut. En utfordring er å lenke geografisk mobile personer, siden flyttelistene i kirkebøkene er så ufullstendige. Ved RHD gjøres det ved å lenke regionvis.

I forbindelse med å lage familierelasjoner, har vi mottatt alle folketellingene fra Digitalarkivet. Dette er et overkommelig datasett, og det dekker hele befolkningen. Bortsett fra emigrerte barn, vil de fleste som har levd frem til voksen alder etter 1801, være med i minst en nominativ folketelling. Vi har derfor laget en algoritme der man går igjennom alle personene i en folketelling og søker i Digitalarkivet med et API-program etter aktuelle personer i hele kommunen. Familier i folketellingene sees samlet slik at man kan sammenligne hele familien med hele familier i andre folketellinger og utvalg av familiemedlemmene i kirkebøker og andre kilder. Søket gjøres først ved et søk på fornavn, etternavn og kommune. Er det få treff, kan søket utvides ved å splitte dobbeltnavn til søk på hvert enkelt navn og deretter utvide søket til hele fylket eller landet. For alle funn sammenlignes likhet i fornavn, etternavn, fødselsår, fødested og bosted med informasjonen i folketellingen. Hvis disse variablene er tilstrekkelig nær, så gjøres et nytt søk på denne hendelsen i en folketelling, kirkebok eller annen kilde. Da gjøres en sammenligning av blant annet andre familiemedlemmer i samme hendelse. Hvis disse også er tilstrekkelig nær, etableres en lenke. Erfaring med utgangspunkt i 1891-tellingen viser at denne metoden gir 0,3–0,6 lenker per personforekomst i folketellingene på steder der de fleste kirkebøker ikke er transkribert, og 0,6–2,5 lenker per personforekomst i folketellinger der de fleste kirkebøkene er transkribert. Det er vanskelig å estimere feilraten. Når man finner feil, er det ofte lett å fjerne feil lenker manuelt, men å kontrollere hele datasettet blir en stor jobb. For Kristiania tilpasses metoden ved å kreve eksakt samme fødselsår i personforekomstene. Dette er nødvendig for å holde feilraten lav, men forhindrer lenking av personforekomster med feil i fødselsår. I de fleste kommunene er over 95 % av lenkene som legges inn med denne metoden, ikke etablert tidligere.

Dette er en fleksibel lenkingsstrategi. Hvis det legges ut en ny kilde i Digitalarkivet, kan man søke ut fra en folketelling som gir stort overlapp med den nye kilden. Nye manuelle lenker kan i noen tilfeller være grunnlaget for nye maskinelle lenker i andre kilder.

Oppsummering

Historisk befolkningsregister gjør det mulig å følge stadig større deler av den norske befolkningen med vel 9 millioner personer fra 1801 til Det sentrale folkeregisteret overtok fra

1964. Det er opprettet ved å lenke over 8 millioner innførsler om personene i folketellinger og kirkebøker og gjøre slektskap mellom familiemedlemmer eksplisitte som koblinger. Det store kildegrunnlaget, snart nær komplette kirkebøker og folketellinger i perioden 1800–1960 som utgjør ca 80 millioner personforekomster, nødvendiggjør bruk av maskinelle metoder. I tillegg kommer andre kilder med informasjon om emigranter, skole, helse, arveoppgjør, fengslinger mm. Dekningen av både kirkebøker og folketellinger over hele landet og i hele perioden gir mulighet for et vesentlig mer komplett register og med mindre feil i løpet av få år enn det er i dag.

Historisk befolkningsregister åpner helt nye forskningsmuligheter innen historie, medisin og samfunnsfag ved å følge en befolkning på 14 mill. gjennom 7 generasjoner. Forskning på sosial ulikhet, migrasjon og genteknologi og mange andre fagområder får helt nye muligheter. Få andre land har kildegrunnlag for å gjøre dette. Norske arkiver vil få en mye større betydning for forskning enn de har hatt så langt ved at dataene blir søkbare og koblet mellom kildene på en helt ny måte. Dette er en av de få forskningsinfrastrukturer som bare øker i verdi med tiden fordi det gir mulighet til å forske på nye problemstillinger, ved å stimulere til ytterligere transkribering av flere historiske kilder og ved å utvide med nye generasjoner.

Historisk befolkningsregister kan potensielt fullstendig endre slektsforskning og skrivning av bygdebøker og annen lokalhistorie. Det vil være lettere å starte på slektsforskning og lettere å få oversikt over familier og lokalmiljø. Man kan i større grad bygge på andres arbeid og utveksle informasjon med eksperter. Slektsforskere har fått et verktøy slik at egne funn kan formidles til andre og bevares for ettertiden. Skoler har fått et nytt verktøy i formidling av lokalhistorie. Siden befolkningsregisteret er basert direkte på søk i kildene og alle hendelsen i kirkebøker vil bli forsøkt lenket til personer i folketellingene, vil det sikkert bli noen overraskende funn slektsforskerne ikke tidligere har funnet. Det vil gi et mer utfyllende bilde av historien til personer og familier. Ofte viser det seg av historien var mer komplisert enn det bildet som familietradisjonen tegnet senere. På den annen side vil slektsforskning ha de samme utfordringer og gleder det alltid har hatt selv om man har fått et nytt verktøy. Det vil alltid være mulig å finne mer informasjon om personene man er interessert i, og man må alltid vise den samme kildekritikk. Den ivrige slektsforsker kan komme lengre enn befolkningsregisteret som skal arbeide med hele befolkningen. Befolkningsregisteret går i liten grad tilbake til før 1801 som kanskje er den perioden de fleste slektsforskere bruker mest tid på.

Artikkelen forklarer hvordan dataprogrammer brukes for å lenke personinnførsler fra ulike kilder og koble opplysninger om familieforhold. Lenkingsalgoritmene er basert på likheter i navn, fødselsår, fødested, yrke, adresse og forhold til familiemedlemmer. Det er delte meninger blant forskere om hvorvidt ikke-stabile egenskaper som yrke og adresse kan brukes som lenkingskriterier. Vi har valgt å anvende disse, men merker slike lenker for at forskere som er skeptiske, kan filtrere bort disse. Spesielt nyttig er det å lenke personpar som hørte sammen, fordi det gir større sikkerhet for riktige lenker basert på to personers egenskaper. Vi har også eksperimentert med lenking av personer med sjeldne navn og kilder som inneholder fødselsdatoer, hvor det er tatt høyde for manglende nøyaktighet. Manuelt genererte lenker er vanskeligere å dokumentere, men de kan i det minste merkes som sådanne. I tillegg sjekker vi at hendelsene utgjør sannsynlige livsløp. Algoritmene må tilpasses egenskapene til hver kilde for å holde feilratene lave. For eksempel er det nødvendig å kreve sterkere likhet i navn og andre egenskaper for store kommuner som Oslo enn i mindre kommuner hvor færre hadde samme fødested. Slike forhold gjør at lenkede datasett ikke er like statistisk representative for hele befolkningen som folketellingene er. Vi håper imidlertid å få finansiert utvikling av teknikker som justerer for disse avvikene.

Datamaskinalgoritmer for lenking og kobling av familieforhold er laget både ved UiT – Norges arktiske universitet og Norsk Regnesentral, og lenkene utveksles mellom partnerne. På nettsted histreg.no kan slektsforskere opprette og endre lenker og familieforhold manuelt, til felles nytte for både genealoger og forskere. Registreringsentral for historiske data ved UiT var pionerer innen lenking og kan levere lenkede data i form av standardiserte filer. Det kan søkes i lenkede datasett på nettstedene rhd.uit.no ved UiT og histreg.no driftet av Arkivverket. Hvordan de ulike inngangene til Historisk befolkningsregister allerede kan benyttes, beskrives i herværende artikkel om brukergrensesnitt.¹⁰

Forfatterne takker de andre deltakerne i prosjekt blant annet Kåre Bævre, FHI, Lars Nygaard, Arkivverket, Jørgen Modalsli, tidl. SSB, Hilde Leikny Sommerseth og Trygve Andersen, UiT – Norges arktiske universitet og Chris Nyborg Nasjonalbiblioteket.

Referanser

- Bouchard, G. (1992). Current Issues and New Prospects for Computerized Record Linkage in the Province of Québec. *Historical Methods*, 67–73. <https://doi.org/10.1080/01615440.1992.9956344>
- Christen, P. (2012) Data Matching. ISBN 978-3-642-31164-2
- Goeken, R., Huynh, L., Lenius, T., Vick, R. (2011). New Methods of Census Record Linking. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 44(1), 7–14. <https://dx.doi.org/10.1080/01615440.2010.517152>
- Holden, L. & Boudko, S. (2018). The Norwegian historic population register and migration. *Journal of Migration History*, 4(2), 249–263. <https://doi.org/10.1163/23519924-00402002>
- Fure, E. (2004). – *en besynderlig Regelmæssighed: dødeligheten i Asker og Bærum på 1800-tallet med særlig vekt på spedbarnsdødeligheten* (Doktoravhandling). Universitetet i Oslo.
- Holden, L., Thorvaldsen, G. & Bråthen, T. R. (2012). Historisk befolkningsregister og DNF 1814. *Heimen*, 49, 399–414.
- Kjelland, A. (2018). Mapping and Analysing Remigration Based upon Norwegian Farm- and Genealogical History Projects. *Journal of Migration History*, 4(2), 314–329. <https://doi.org/10.1163/23519924-00402005>
- Lokalhistoriewiki, (2020) .Bygdebøker. Hentet fra <https://lokalhistoriewiki.no/index.php/Bygdebok>.
- Ruggles, S., Fitch, C. & Roberts, E. (2017). *Historical Census Record Linkage*. Working paper, Minnesota Population Center, University of Minnesota, 2017–3.
- Thorvaldsen, G. (2006). *Away on Census Day. Enumerating the Temporarily Present or Absent*. *Historical Methods*, 39(2), 82–96. <https://doi.org/10.3200/HMTS.39.2.82-96>
- Thorvaldsen, G. (2011). Using NAPP Census Data to Construct the Historical Population Register for Norway. *Historical Methods*, 44(1), 37–47. <https://doi.org/10.1080/01615440.2010.517470>
- Thorvaldsen, G., Andersen, T. & Sommerseth, H. L. (2015). Record Linkage in the Historical Population Register for Norway. Red. I G. Bloothoof, P. Christen, K. Mandemakers & M. Schraagen, *Population Reconstruction* (s. 155–170). Heidelberg: Springer.
- Thorvaldsen, G. & Østrem, N. O. (2018). Migration and the Historical Population Register of Norway. *Journal of Migration History*, 4(2), 237–248. <https://doi.org/10.1163/23519924-00402001>
- Thorvaldsen, G., Sommerseth, H. & Holden, L. (2020). Anvendelser av Norges historiske og moderne befolkningsregistre. *Heimen*, s. 230–243. <https://doi.org/10.18261/issn.1894-3195-2020-03-05>
- University of Minnesota, Minnesota Population Center, (2020). <https://ipums.org>.

10. Thorvaldsen et al. 2020.