

# Historisk befolkningsregister og DNF 1814

*Av Lars Holden<sup>1</sup>, Gunnar Thorvaldsen<sup>2</sup> og Torkel Rønold Bråthen<sup>3</sup>*

Norsk Regnesentral, Registreringsentralen for historiske data, Arkivverket

## Innledning

Byggingen av Historisk befolkningsregister (HBR) er nå i full sving, i første omgang for perioden 1801 til 1815. Målsettingen er både å rekonstituere Norges befolkning i dagene rundt 17. mai 1814 og å konstruere en nominativ folketelling for å supplere den statistiske folketellinga fra 30. april 1815. Ved ferdigstillingen til grunnlovsjubileet i 2014 blir dette verdens første allment tilgjengelige og landsdekkende befolkningsregister og dekker perioden 1801–1815.<sup>4</sup> Dette delprosjektet har fått navnet «Det norske folk i 1814» (DNF1814), og initiativet er tatt av Riksarkivaren og realiseres av representanter for Arkivverket, Registreringsentral for historiske data ved Universitetet i Tromsø, Norsk Regnesentral, DIS-Norge og andre frivillige. DNF1814-prosjektet utgjør kun en mindre del av den åpne delen av HBR.<sup>5</sup> HBR tar nemlig mål av seg til å dekke hele den norske befolkningen fram til rundt 1930, og i dette større prosjektet vil også flere partnere delta. At Norge kan ligge langt framme på dette feltet, skyldes at vi sammen med de andre nordiske landene har fullstendige kirkebøker som fyller hullene mellom folketellingene.<sup>6</sup> Det planlegges også en lukket del som knytter an til Det sentrale folkeregister (DSF) som ble opprettet i 1964.<sup>7</sup>

HBR skal knytte samme all åpne historisk

person- og stedsinformasjon om personer i Norge. Det skal bygges hovedsakelig på kirkebøker og folketellinger, og vil etter hvert også inkludere annen informasjon som emigrantlister, skiftebøker, fengselsprotokoller og gravminner. Det skal knyttes sammen informasjon om samme person som forekommer i flere kilder og opplysninger om bosteder. Dette er en enorm oppgave som aldri vil bli fullendt. Perioden fra 1735 til 1964 inneholder 9,7 millioner personer og 37,5 millioner kildehendelser i folketellinger, kirkebøker og andre kilder.<sup>8</sup> HBR kan nettopp oppfattes som et register til alle disse kildene fordi det gir en samlet oversikt over samme person og familie fra flere kilder og ulike bosetninger.

## Nye muligheter med HBR

DNF 1814 og senere utvidelser av HBR vil kunne brukes i en rekke forskjellige lokale, regionale og nasjonale studier og gi grunnlag for internasjonal komparasjon. HBR vil blant annet inngå i et internasjonalt samarbeid gjennom NAPP.<sup>9</sup> Det vil gi oss en ny historisk og samfunnsvitenskapelig innsikt i de aktuelle periodene. Med longitudinelle mikrodata kan man studere blant annet hvordan familiestruktur og sosial- og geografisk mobilitet endret seg kontinuerlig i motsetning til øyeblikksbilder gitt i folketellingene. I et medisinsk perspektiv

vil HBR være en viktig kilde for studier av blant annet genetiske sykdommer,<sup>10</sup> genene til den norske befolkning,<sup>11</sup> og studier av inngifte.<sup>12</sup> Bibliografiene ved Demografiska databasen i Umeå<sup>13</sup> og Minnesota Population Center i Minneapolis<sup>14</sup> inneholder svært mange forskningstemaer som ikke er studert for Norges vedkommende, så mulighetene er mange.

En sentral demografisk problemstilling gjelder utviklingen av folketallet fra 1801 til 1825. I *Norsk innvandringshistorie*<sup>15</sup> påstås en utvandring på to prosent fram til 1815 og deretter en tilsvarende innvandring fram til folketellinga i 1825. Tidligere har Michael Drake vært enig med både Eilert Sundt og Gunnar Jahn om at dette skyldes 1815-tellingas lave kvalitet med en undertelling på ca to prosent.<sup>16</sup> Det blir spennende å se i hvilken grad DNF 1814-prosjektet kan støtte innvandringshistoriens spenstige forsøk på å opprette tillit til 1815-tellinga – for øvrig uten å diskutere den nevnte historiografien.

Innen lokalhistorie og slektsforskning blir det lettere å sette egen familie- og samfunns-historie inn i en større sammenheng. Rent praktisk vil det bli lettere å finne fram til kildene og mer rasjonelt å knytte an til andres arbeid. Det vil alltid være mulig å finne mer informasjon, flere kilder og sammenlignbare historier. Det vil i mindre grad bli nødvendig å gjenfinne de samme individene og dublere de samme lenkene, og vi kan i stedet bygge videre på andres arbeid og komplettere kildene for hver enkelt person og slekt. Resultatet blir en database hvor vi kan følge individer, familier, gårder, plasser og andre lokaliteter over tid. Utfordringene gjennom hele tidsperioden er de samme. Det er krevende å etablere et slikt steds- og befolkningsregister, og denne artikkelen beskriver noen av de utfordringene vi møter i arbeidet og hvordan de løses. Spesielt gjelder dette lenkingen som skal skje gjennom en

større dugnad på Internett. Målet er realistisk hvis vi løfter i flokk. HBR representerer ny teknologi for å samarbeide om lenking av personopplysninger. Dette er beskrevet i avsnittet om lenking og bygger på programvaren Mediawiki, en variant av systemet som brukes i Wikipedia.<sup>17</sup> Framstillingen inneholder en del metodiske og tekniske detaljer som ikke er nødvendige for bruken av systemet. De lesere som er mest interessert i utnyttelsen av de historiske kildene, kan altså springe videre om noen av avsnittene oppleves som tekniske. Først må vi gjennomgå hovedprinsippene i oppbyggingen av HBR.

## Hovedprinsipper

HBR vil hovedsakelig basere seg på innlesning av data fra kilder, i første omgang folketellinger og kirkebøker. Det sikrer at vi inkluderer alle personpostene<sup>18</sup> i disse kildene én, og bare én gang. Utfordringen blir å knytte sammen personene og stedene mellom de forskjellige kildene. Dette vil skje dels automatisk og dels manuelt. Hovedprinsippene i HBR er følgende:

1. HBR skal bygge på flest mulig kilder av god kvalitet. Det skal være med toveis pekere, det vil si fra databasen til kildene og vice versa.

2. Vi ønsker størst mulig åpenhet og transparens. Dermed blir det mulig å se hvem som har gjort endringer og begrunnet disse.

3. Vi vil motivere til størst mulig bruk og flest mulig bidrag til utviklingen av god kvalitet. Alle brukere får anledning til å kommentere databasens kvalitet.

Disse prinsippene sikrer entydighet og referansene til kildene. Det skal sikre mest mulig data av god kvalitet og sikre at kvaliteten øker over tid. HBR har funksjon som et indeks eller register til kildene istedenfor å erstatte disse. Både med hensyn til kildereferanser og åpenhet, vil HBR være ulik det historiske befolkningsregistret i Island, *deCode*. Dette systemet

er lukket slik at folk kun har tilgang til egne aner, og det er heller ikke systematiske referanser til kildene eller begrunnet hvordan personpostene er lenket.<sup>19</sup> De historiske, longitudinelle databasene i Sverige og Nederland har god tilknytning til originalkildene, men dekker bare deler av disse landene. *European Science Foundation* støtter et nettverk som arbeider for å opprette tilvarende historiske befolkningsregistre i andre europeiske land.<sup>20</sup>

## Bruk av kilder

I Norge har det vært ført kirkebøker fra 1623 med innførsler om døpte, viede og døde. Føringen har vært påbudt siden 1680. Vi regner imidlertid med en manko på 10–20% på 1700-tallet.<sup>21</sup> Det er noe av bakgrunnen for at HBR starter i 1801 med folketellinga samme år som sitt utgangspunkt, men det skyldes òg at de tidligste kildeseriene som oftest er informasjonfattige ved at personenes alder kan mangle, navn på gifte kvinner utelatt i flere hendelser og bosted ikke oppgitt.

I Digitalarkivet etableres nå entydige identiteter (IDer) for hver person- og eiendomsforekomst i kildene som transkriberes, henholdsvis PKID og EKID. Disse IDene brukes som referanser i HBR. Det noe komplekse formatet på disse maskinlesbare referansene bør ikke bekymre brukerne. Når referansen er tilgjengelig med et klikk, blir det lett å undersøke om det finnes ytterligere eller motstridende informasjon tilgjengelig fra samme kilde. Kombinasjon av flest mulig kilder av god kvalitet vil forbedre HBR både ved at det gis mer informasjon om personer og steder, og ved at lenkingen blir sikrere. Komplementære kilder kan være alt fra skifteprotokoller til avisoppslag. I utgangspunktet leses hele kilden inn i HBR og det etableres en side for hver eneste person- og eiendomsforekomst i kilden med toveis pekere mellom de forskjellige personene og stedene

nevnt i kilden og HBR. Det sikrer at man lett kan bevege seg mellom én person eller ett sted nevnt i HBR og kilden der personen eller stedet er nevnt. Noen kilder egner seg ikke for å bli innlest i sin helhet, eksempelvis listene fra Ellis Island over immigranter til USA og Danmarks Adels Aarbok, der bare en mindre andel av personforekomstene er relevante for Norge. Da etableres istedenfor en kildetabell i HBR med lenker både til personsider i HBR og stedet der personen omtales i kilden. Da er det lett å holde oversikt over hvilke personer som er registrert i HBR og lett å legge til flere personer i HBR som omtales i kilden. En kildetabell i HBR vil sikre entydighet.

## Bearbeiding av dataene

Folketellingene inneholder informasjon om stilling i familie og hushold, som gir grunnlag for å gruppere individene i familier. HBR etablerer familiene automatisk ved innlesning av folketellinger basert på informasjon om familiestilling, rekkefølge i bostedslisten, kjønn, alder og navn. Automatikken bygger på et dataprogram utviklet av våre partnere ved Minnesota Population Center som en del av *North Atlantic Population Project* (NAPP).<sup>22</sup>

I alle referanser til kildene brukes en kildetro gjengivelse av navn, fødselsår med videre. Samme person vil dermed kunne ha variantstavinger av navn, avvikende alder og lignende fra forskjellige kilder. Til den overordnede presentasjonen av personen velger vi imidlertid ut de kjernedata vi har mest tiltro til. De velges ut fra følgende prioriteringsrekkefølge: 1) manuell overstyring, 2) kirkebøker der eldre (samtidige) går foran nyere, 3) folketellinger der nyere går foran eldre og 4) andre kilder der nyere går foran eldre.

## Lenking og kobling

Lenking kan defineres som det å identifisere

og fastslå at samme person er omtalt i minst to forskjellige kilder. Kobling er å knytte sammen forskjellige familiemedlemmer og baserer seg på kilder som nevner flere personer i samme hendelse. Ulike lenkingssystemer er beskrevet i *Heimen* tidligere.<sup>23</sup> Her repeterer vi kort at familierekonstruksjonsmetoden kan føres tilbake til Louis Henry på 1950-tallet, og ble videreutviklet av Ståle Dyrvik, Sølvi Sogner og Lajos Juhasz på 1960- og 1970-tallet. Metoden er utviklet med bruk av datamaskin i Etne- og Rendalsdatabasene. Lars Nygaard og Eli Fure anvendte en mer generell programvare for å lage et befolkningsregister for Asker og Bærum. Programvaren «Busetnadsog» fra Arnfinn Kjelland og Ole Martin Sørungård bruker bosted som viktig lenkingspunkt, og databasen brukes til å lage gårds- og slektshistorie. Automatisk søking og lenking ble programmert for å lenke deler av befolkningen i Troms på 1800-tallet.

Flere genealogiske databaser bruker nå programvaren Mediawiki på Internett.<sup>24</sup> *WeRelate.com* er størst med over 2,1 millioner personer registrert i databasen. *Geni* er en tilsvarende database som ikke bruker Mediawiki.<sup>25</sup> Dette er imponerende systemer med mye grafikk og vel lagt til rette for mange brukere. Disse er i stor grad basert på å lese inn ferdige slektstrær via gedcom-filer fra slektsforskere. Siden de ikke tar direkte utgangspunkt i kildematerialet, får de dermed problemer med dubletter og kvalitet i databasen.<sup>26</sup>

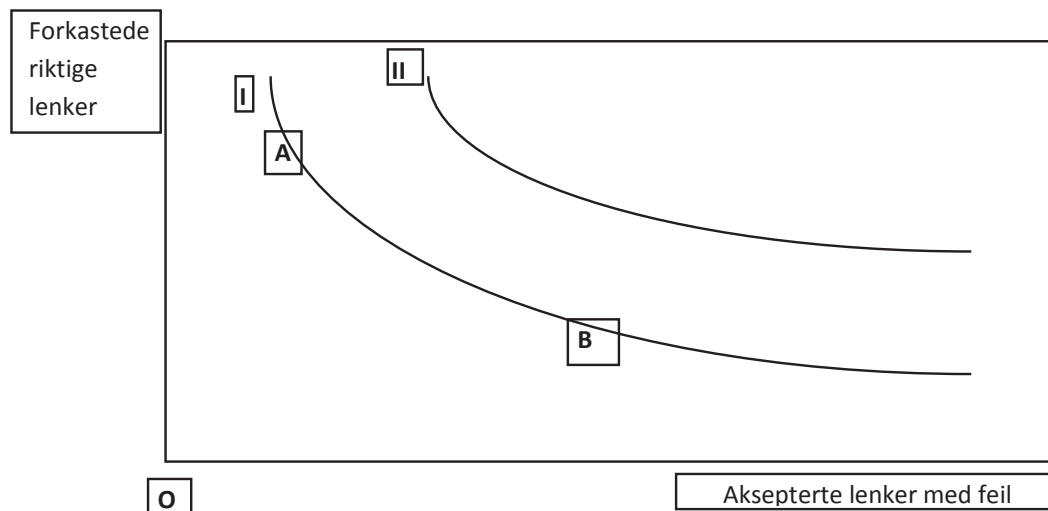
HBR representerer en ny teknikk for å lage befolkningsregistre. Størrelsen på prosjektet gjør dette nødvendig. I HBR leses alle hendelsene direkte fra de viktigste kilder og lagres som egne sider. Det setter store krav til at systemet er oversiktlig, enkelt, transparent og utviklet slik at kvaliteten styrkes over tid istedenfor å degenereres. En sentral del i lenkingsmetodikken blir dermed hvordan den kan motivere

mange brukere til å gi bidrag av god kvalitet og hvordan disse bidragene følges opp og kvalitetssikres. Vi har valgt wikiteknologien fra Mediawiki, som løser disse utfordringene på en god måte. For å holde orden på alle de forskjellige gruppene av personer brukes *kategori*-begrepet i Mediawiki aktivt. Det gjør at det er lett å finne alle personer omtalt i en bestemt *kilde*, alle *familier i en kommune*, alle *prester i Norge* eller andre grupperinger av personer, familier og steder.

Det nasjonale siktemålet gjør det mulig å fange opp migrasjon mellom forskjellige områder og til/fra utlandet. Personer som flytter mellom områder, vet vi skal være mulig å gjenfinne i begge områdene, og det etableres spesielle lister for migrerte personer som ikke er gjenfunnet. Tilsvarende vil det være mulig å gjenfinne røttene til personer av norsk avstamning bosatt i utlandet i HBR. Fra et forskningssynspunkt vil det være spesielt interessant å følge de personene som flytter og som vi i mindre grad har fanget opp i andre databaser. Mange vil sikkert like utfordringen med å lenke personer som har flyttet. Erfaringene fra andre wikier, som Lokalhistoriewiki,<sup>27</sup> viser at forskjellige bidragsytere vil dele kunnskap ut fra sin spesielle kompetanse. Temaene kan for eksempel være egen slekt og bosted, geografiske områder, bestemte yrker, etniske grupper, kjente/omtalte personer fra andre kilder og sekundærlitteratur, standardisering av navn og retting av skrivefeil. Det er mulig å beskrive arbeid innen et bestemt tema i form av prosjektsider. Vi regner med at det vil gi mange brukere med forskjellige innfallsvinkler til de samme dataene. Dette vil igjen heve kvaliteten. De forskjellige bidragsyterne vil bruke forskjellige teknikker.

For alle lenker etableres en indikator for kvaliteten av lenken på en skala fra 0–10. Som en retningslinje brukes følgende skala:

10: helt sikker lenke for eksempel ved at



Figur 1. Forholdet mellom aksepterte lenker med feil og riktige lenker som ble forkastet i kilder fra tidlig 1800-tall (kurve II) og seint 1800-tall (kurve I). Etter Johansen 2002.

man kjenner personen via familie eller som kjent person.

8: samme fødselsdato og samme eller lignende navn samt geografisk tilknytning.

6: samme eller lignende navn på begge ektefeller og med geografisk tilknytning.

4: samme navn, passende alder og knyttet til samme sted.

1: regner under tvil som sannsynlig lenke, gir begrunnelse for lenke.

0: etableres som lenkingskandidat. Gir begrunnelse.

Hvis en bruker er usikker på om det er samme person, kan altså de to personforekomstene registreres som lenkingskandidater. Da slås ikke sidene sammen, men det etableres pekere mellom sidene.

### Automatisk eller manuell lenking?

Bare i DNF1814-delen av HBR-prosjektet skal informasjon om rundt en og en kvart million innbyggere lenkes, det vil si omtrent 888 000 i 1801-tellinga, de 350 000 fødte fram til

1815 pluss et ukjent, men relativt lite antall immigranter til Norge. Dette krever betydelige ressurser og gjør det interessant å vurdere hvor stor del av lenkingsjobben som lar seg automatisere. Det er vel verdt å merke seg hva pioneren Hans Chr. Johansen har skrevet om dette – de danske kildene samsvarer jo med de norske i dette tidsrommet.<sup>28</sup>

Med referanse til figur 1 er det to målsettinger, for det ene å lenke korrekt så mange individposter som mulig og for det andre å unngå å lenke innførsler som ikke hører sammen. Ideallet befinner seg i origo, nederst til venstre, der alle lenkede innførsler egentlig hører sammen og ingen lenker er uteglemt. Problemet er at når vi stiller strengere krav til lenkingsreglene for å redusere risikoen for å lenke sammen poster fra ulike individer (horisontal akse), vil vi samtidig utelukke flere lenker som egentlig burde vært med (vertikal akse). Og omvendt: Dersom vi innfører mer liberale lenkingsregler for å sikre at alle potensielle lenker realiseres (langt nede på vertikal akse), vil vi komme til å inkludere

flere lenker som egentlig skulle forkastes (når vi går mot høyre på horisontal akse).

Hans Chr. Johansen mener at akseptable lenkingsresultater ligger mellom punktene A og B på kurve I i diagrammet, enten med få aksepterte feilaktige lenker (A) eller få utelatte riktige lenker (B). Lenkingsresultat langs kurve II mente han var uakseptable fordi det både ble innført mange feilaktige lenker i databasen og mange lenker var utelatt for personforekomster som burde forenes. Og så hans helt sentrale poeng i vår sammenheng: Fra og med siste halvdel av 1800-tallet har vi kildemateriale med en presisjon som tillater automatisk lenking langs kurve I, mens kilder fra tidligere tidspunkt er så upresise at de må lenkes manuelt, ellers havner vi på kurve II. Heldigvis har vi gjennom det siste tiåret utviklet teknikker og metoder som tillater oss å konkludere noe annerledes, i hvert fall for deler av materialet.

For det første kan vi nå på en fleksibel måte føre sammen variantstavinger av samme navn både ved hjelp av standardiserte navnelister og algoritmer for sammenligning av tekststrenger uten at presisjonen i navnene blir nevneverdig mindre. For det andre har vi utviklet programvare som ikke bare lenker enkeltindivider, men tar i betraktning (ekte)par eller hele familier i lenkingen. Begge deler er særlig verdifullt i en periode der andre opplysninger enn personnavn (alder, fødested) er upresise eller mangler helt i kildene. Ved å utnytte dette kan vi flytte den delen av befolkningen som lar seg identifisere sammen med andre i flere kilder, fra kurve II til kurve I i figur 1. Det er en betydelig andel, og en spesiell databehandling av 1801-tellinga viser at omtrent 80 % av kvinnene og 85% av mennene bodde sammen med minst ett annet familiemedlem.<sup>29</sup>

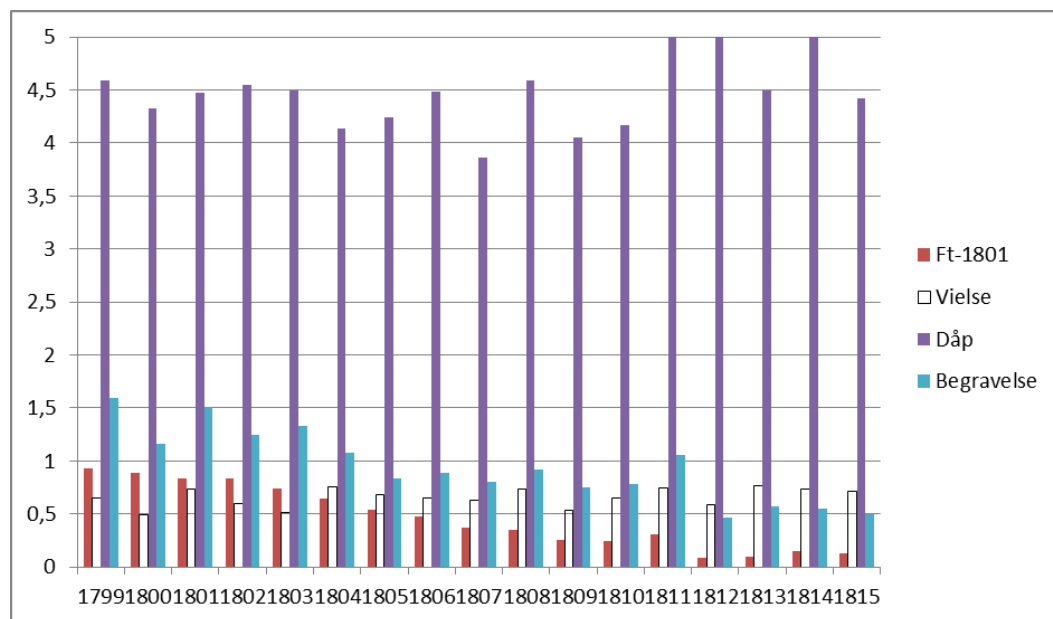
Den prinsipielle problemstillingen med avveiningen mellom antall lenker og andelen ukorrekte lenker reduseres også i HBR ved

muligheten for manuell lenking. Samme krav for lenking gjelder i hele databasen og det gis en kvalitetsindikator og begrunnelse for hver lenke slik at hver enkelt bruker kan gjøre en selvstendig vurdering og velge sin avveining. I tillegg kommer muligheten til å etablere lenkingskandidater og overprøvingen av andre brukere.

## Automatisk lenking i Lenvik

Datakonsulent Trygve Andersen ved Registeringssentralen har utviklet programvare for automatisk lenking av ektepar og andre familie-medlemmer som utnytter spesielle algoritmer for sammenligning av navn.<sup>30</sup> I forbindelse med DNF1814-prosjektet kjører vi nå disse programmene mot 1801-tellinga og kirkebøkene for Troms som er transkribert for den aktuelle perioden. Utgangspunktet er de menn som var fedre i en eller flere av de 696 dåpshandlingene i tidsrommet 1799 til 1815, fra 16 (1811) til 63 (1806) årlig. Histogrammet i figur 2 viser hvor mange andre kildeposter det var mulig å lenke til i form av gjennomsnitt for fedre hvert år. Automatisk lenking gir best resultat internt i dåpslistene der hvor både far og mor er ført sammen år etter år. Rundt to tredjedeler av foreldrene er lenket til vielsen, et stabilt tall over tid. Begge ektefellenes navn finnes i begge kirkebokslistene og prestene hadde gjerne samme skrivemåte. Det største problemet er lenking til 1801-tellinga, for her vil de fleste par være ført på ulike bosteder før de giftet seg og fikk barn sammen. Dermed får vi gode lenkingsresultat tidlig i perioden, og langt svakere mot slutten fordi en større andel av parene har giftet seg etter 1801. De begravelsene som ble lenket til far ved dåp er stort sett barn; resultatet både med hensyn til lenking av dåp og begravelse må altså forstås på bakgrunn av den høye fertiliteten og barnedødeligheten i perioden.

Hvis bostedet er oppgitt i kirkeboka, og er



Figur 2. Gjennomsnittlig antall lenker fra far ved dåp til dåps-, vielses- og begravellesliste i perioden 1799–1815 samt 1801-tellinga for Lenvik.

det samme som i 1801, blir det også mulig å lenke enkeltindivider uten oppgitt alder i kirkeboka. I perioden er det 834 begravelser. Her er det 349 begravelser som ikke er lenket til noe når det tas utgangspunktet i den avdøde. Især for voksne menn er det vanlig at de føres inn i kirkeboka uten referanse til slektninger. Også lenger ut i hundreåret blir mange stående ulenket, det gjelder minst en tredjedel av personene som ble forsøkt lenket mellom folketellingene 1865 og 1875 for deler av Troms.<sup>31</sup> Vi ser med andre ord at det er betydelig rom for manuell lenking ut fra den detaljerte kunnskap om slekter og lokalsamfunn som heldigvis fins i rikt monn i Norge. Med HBRs wikisystem på Internett kan vi altså samle denne kunnskapen i en stor dugnad og løse mange av de nøttene vi som enkeltgranskere sliter med.

## Oppbygging av wikien

HBR bygges som en wiki på linje med Wiki-

pedia. Den bruker programvaren Mediawiki. Wikien er bygget over en MySQL database og programmeres i språkene PHP og C. I en wikidatabase lagres alle endringer og «stemples» med tid for endringen og hvem som utfører den. Det er lett å reversere endringen. For hver side i en wiki er det en diskusjonsside der man kan drøfte innholdet på siden, for eksempel diskutere om en lenke er riktig eller ikke.

Det etableres sider for hver enkelt person og for hver familie. En personside har navn av typen «Ole Hansen (21)», som forteller at det er den 21. personforekomsten med navnet Ole Hansen som er registrert i HBR. Figur 3 viser et eksempel på en *personsida*. Her kan man samle alle kildeopplysninger om personen, lenker til familiemedlemmer og ha mulighet til å legge inn bilde og skrive biografi. En *familiesida* får navn av typen «Ole Hansen (21) og Johanne Thorsdatter (8)». Bosteder opprettes fra folketellingene og har pekere til beboere ved hver fol-

## Axel Smith (1)

**Fra Historisk befolkningsregister**

Axel Smith, mann  
 PersonID: pd01003108010772  
 Født : Ukjent dato  
 Død : Ukjent dato

### Partner og barn

Partner: Fredrikke Debora f. Sunde  
 Barn 1: Ragnhild Axelsdatter  
 Barn 2: Sverre Axelsen  
 Barn 3: Georg Axelsen  
 Barn 4: Erik Axelsen

### Livsløp

Nr	Dato	Kilde	Rolle	Navn	F.dato	F.sted	Bosted	Famst.	Sivst. Yrke	PKID	Far/Brudgom	Mor/Brud	Partner	Barn/Andre
1	1876.04.17	Kirkebok, dåp	far	Axel Smith	1849		Ø. Slotsg. 12 b	Fuldmægtig			DA		Fredrikke Debora f. Sunde	Ragnhild Axelsdatter
2	1877.03.25	Kirkebok, dåp	far	Axel Smith	1848		Ø. Slotsg. 12 b	Handelsfuldmægtig			DA		Fredrikke Debora f. Sunde	Sverre Axelsen
3	1878.07.19	Kirkebok, dåp	far	Axel Smith	1849		Thurins Gd i Ø. Slotsg.	Kontorchef			DA		Fredrikke Deborah f. Sunde	Georg Axelsen
4	1881.12.09	Kirkebok, dåp	far	Axel Smith	1819		Ø. Slotsgade 12 B	Kjøbmand			DA		Fredrikke Deborah f. Sunde	Erik Axelsen

Linjenr.: (1) (2), 6, Mon, 02 Jan 12 21:08:52 +0100, Lars Holden, Samme navn på person og ektefelle. Samme adresse.  
 Linjenr.: (1, 2) (3), 6, Mon, 02 Jan 12 21:10:06 +0100, Lars Holden, Samme navn på person og nesten samme på ektefelle.  
 Linjenr.: (1, 2, 3) (4), 6, Mon, 02 Jan 12 21:11:16 +0100, Lars Holden, Samme navn på person og ektefelle. Samme adresse.

### Lenkingskandidater

Ingen

### Steder

### Biografi

Hentet fra «[- Denne siden ble sist endret 16. jan 2012 kl. 08:15.
- Denne siden er vist 25 ganger.
- Personvern
- Om Historisk befolkningsregister
- Forbehold](http://slekt.nr.no/wiki/Axel_Smith_(1)»</a><br/>
    Kategorier: Oslo, Johannes menighet, dåp, far 1875-1881 | Oslo, Johannes menighet, personer 1850-1899
  </p>
<ul style=)

Figur 3. En personside i HBR.

ketelling. Navnet er av typen «Møre og Romsdal, Surendalen, Bakken (3)». Nummer brukes bare for bosteder der det er nødvendig for å gi entydighet. Bosteder etableres ikke fra kirkebøker fordi de ikke er presise nok i beskrivelse av stedene. Person-, familie- og stedssidene opprettes ved innlesning fra hver enkelt kilde (i første omgang folketellinger og kirkebøker). Sidene slås sammen når vi har tilstrekkelig sikkerhet for at det er samme person, familie eller

bosted som omtales i forskjellige kilder.

Siden viser en far som er lenket mellom 4 personforekomster i dåp. Vi ser at alder og navn på ektefelle varierer noe mellom personforekomstene. I andre linje i livsløpstabellen står «Linjenr.: (1,2)(3), 6». Det viser at sammenslåingen er utført ved å slå sammen linje 1 og 2 med linje 3 i livsløpstabellen og denne lenkingen er gitt 6 i kvalitet. Kategoriene nederst viser noen av kildene. Siden kan testes interaktivt på [www](http://www.www).





som har store lenkede samlinger av den norske befolkningen. Det vil også være automatiske og halvautomatiske lenkingsmetoder tilgjengelig i HBR. Det sikrer at så snart en ny kilde er lest inn i HBR, kan de letteste lenkene bli etablert.

Én personside i HBR kan ikke ha flere fedre eller mødre. Det er dermed ikke mulig å slå sammen personsider der det er konflikt mellom foreldrene på de to sidene. Hvis man har to hele slektstrær som representerer de samme personene, må man derfor slå sammen slektstrærne ved å starte med eldste generasjonen og ta én og én generasjon. Når HBR leser inn en lenket samling leses én og én gruppe med personforekomster som skal lenkes. Hvis noen personforekomster ikke kan lenkes på grunn av konflikt mellom lenkene til far eller mor, legges denne gruppen med personforekomster til slutt i listen. Det er jo alltid håp om at foreldrenes personforekomster kan lenkes sammen på et senere tidspunkt. Hvis foreldrenes ulike personsider ikke kan slås sammen i denne omgang, etableres barnas personsider som lenkingskandidater til hverandre.

Listen med PKIDer kan eksempelvis lages med et automatisk lenkingsprogram som sammenligner navn, alder og bosted og beregner sannsynlighet for at det er samme person slik det er gjort i Troms. Men listen kan også lages ved å lese inn et befolkningsregister fra et område, en slekt eller annen tematisk samling. Gedcom-filer kan brukes, men det kreves at kildereferansene eksisterer. Det må gjøres en kvalitetsvurdering før man leser inn lenkede samlinger.

## Manuell lenking

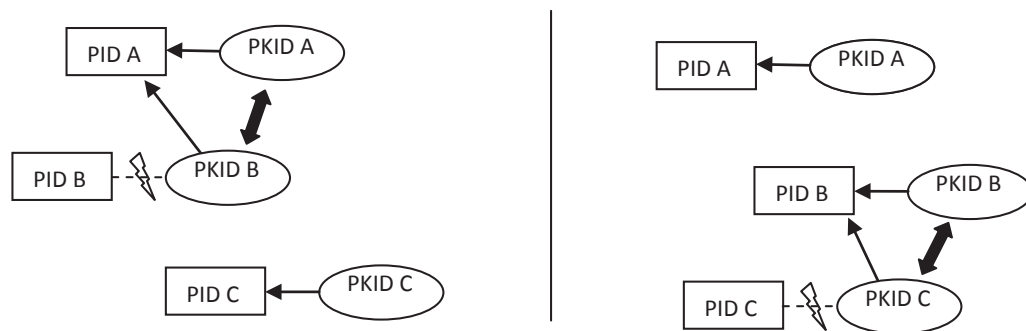
Vi forventer at den dominerende delen av lenkingen og koblingen, især i DNF1814, vil skje manuelt med et bidrag fra mange slektsforskere og interesserte i lokalhistorie i denne store dugnaden. Den manuelle lenkingen kan

skje ved at man gjenfinder personer i HBR og sjekker referanser til kilder mot sin personlige database eller slektsbok. Men den kan også skje ved å søke fram personforekomster i HBR ved å følge kronologisk alle hendelser i en kirkebok, alle hendelser på en gård over tid eller søke ved hjelp av andre kriterier i kildene. Er man i tvil, kan man «prøvelenke» to personforekomster som lenkingskandidater og deretter egenhendig eller ved hjelp av andre bekrefte eller avkrefte lenken. Især i de tidlige kirkebøkene og for emigranter og immigranter kan det være aktuelt å etablere mange mulige lenkingskandidater før man med rimelig sikkerhet kan identifisere en av dem.

Ved sammenslåing av sider oppdateres alle pekere på relaterte person-, familie- og stedssider til navnet på siden for den personforekomsten som er foretrukket og skal brukes. I alle tekster som viser til kilder, beholdes kildetro gjengivelse av navn, datoer og lignende. Hvis det skal lenkes mellom to personforekomster og bare den ene personforekomsten allerede er opprettet i HBR, registreres den andre personforekomsten på den ene siden og de nødvendige lenker etableres. I HBR registreres koblinger, altså det at to personer er relatert til hverandre, eksempelvis foreldre og barn, ved å etablere pekere mellom person- og familiesidene.

## IDer for personer

Det er nødvendig å etablere IDer til alle personer i HBR slik at vi kan ha en sikker referanse til personer internt i HBR, og mellom HBR og eksterne programmer. Dagens personnummer baserer seg på sikker identifisering av fødselsdato. Fødselsdato vil ofte ikke være kjent for historiske personer, og kan derfor ikke brukes som identifikator. Som tidligere nevnt, etablerer Arkivverket IDer for alle personforekomster (PKID) og eiendommer (EKID) nevnt i kilder. Vi regner dette som en sikker identifisering av



Figur 5. Endring av personID (PID) når lenking mellom to personforekomster (PKID) endres. I eksemplet til venstre har PKID A høyest prioritet, og PKID B «arver» PID etter denne når de antas å vise til identisk person (tykk pil). Når så denne lenken blir brutt men PKID C kan lenkes til PKID B, får PKID B tilbake sin PID fordi PKID C har lavere prioritet av de to. I motsatt fall ville PKID B fått PID fra C.

en personforekomst, og knytter personen til en veldefinert kilde. Hvis man senere kommer til at navn eller annen informasjon om personen er transkribert feil, vil ikke det endre PKIDen.

HBR bruker en regelbasert definisjon av personID –PID – ut fra PKID, det vil si at vi lager en prioritetsregel for PKID. PID er definert som den PKIDen med høyest prioritet av lenkene til en bestemt person. Dette er et system som lett håndterer endringer i lenkingen. Dette kan illustreres ved at PKIDene A og B er lenket, men en tredje PKID C som vi så lenker til B, viser at lenken mellom A og B må forkastes. Da endrer vi bare PIDene for personene fra A(B) og C til A og B(C) der parenteser viser hvilke andre PKIDer som er knyttet til denne PIDen (se figur 5).

I HBR brukes følgende prioriteringsrekkefølge for å lage PID ut fra PKIDer:

1. folketelling 1910
2. begravelser/lister over døde før 1910
3. fødsler etter 1910
4. liste over utvandret før 1910
5. liste over innvandret etter 1910
6. andre folketellinger der nyere folketellinger har høyere prioritet enn eldre.
7. andre oppføringer i kirkebøker der eldre

(samtidige)oppføringer har høyere prioritet enn nyere

8. andre kilder der nyere oppføringer har høyere prioritet enn eldre.

De første fem kategoriene er ikke overlappende, og de tre første er antagelig av relativt god kvalitet. Folketellingen 1910 vil senere bli brukt til å lenke mellom HBR og den lukkede delen av befolkningsregisteret fram til det sentrale folkeregisteret i 1964. På personsiden i HBR vises PIDen til personene. Internt i HBR vil imidlertid URLen eller navnet på siden være mer naturlig å bruke som en identifikator.

## Historiske steder og administrative grenser

Utviklingen av befolkningen er selvsagt nært knyttet opp til utviklingen av bosteder, lokalsamfunn og regioner. Det er en ambisjon i HBR-prosjektet å koble til denne informasjonen, som jo er dynamisk, ved opprettelsen av husmannsplasser, oppsplitting av gårder i flere bruk og ved byenes vekst. Et bosted vil være knyttet til flere kommuner over tid hvis kommunegrensene endres.

Vi kan lese inn 1886- og 1950-matrikkelen

og knytte stedssidene fra folketellingene til disse ved hjelp av gårdsnavn, plassnavn, gatenavn, matrikkelnummer, gårdsnummer, husnummer og gatenummer.<sup>33</sup> Resultater fra prosjektet *Historisk administrative grenser*<sup>34</sup> viser at 80 til 90 % av gårdene kan det knyttes entydige matrikkelnummer til. Både restkategorien på landet, for Finnmark som ikke ble matrikulert i vår åpne periode, og for byene vil kreve omfattende spesialarbeid, især sistnevnte grunnet overgangen fra løpenummer i byen til løpenummer i den enkelte gate. Igjen er vi avhengige av lokalhistorisk ekspertise for å legge puslespillet korrekt. Arne Solli ved Universitetet i Bergen har gjort et imponerende arbeid med person- og stedsinformasjon for byene i BERGIS-prosjektet<sup>35</sup> for perioden 1696–1906.<sup>36</sup>

Det planlegges å gi koordinater til alle steder. Det gir muligheter for å lage mye forskjellig grafikk, for eksempel migrasjonsmønsteret til en person eller familie på bakgrunn av geografiske koordinater.<sup>37</sup>

## Transkribering

Med transkribering mener vi datamaskinell avskrift fra kildene via tastatur, til forskjell fra skanning hvor bare bildet, som ikke er søkbart, blir overført til et digitalt format. Tradisjonelt har dette foregått ved å transkribere alle postene i en kilde systematisk og samlet. Deretter blir det kvalitetssikret ved korrekturlesning foretatt av en annen person. Transkriberingen utføres både i statlige institusjoner og av frivillige. Mange kilder er allerede transkribert på denne måten, men det gjenstår fortsatt store deler av dette arbeidet. I HBR-prosjektet arbeides det derfor med å utvikle mer effektive teknikker for transkribering.<sup>38</sup> Kirkebøkene fram til rundt 1930 er tilgjengelige i Digitalarkivet som skannede bilder, noe finnes på mikrofilm, men fortsatt er mye kun tilgjengelig på papir, for eksempel de klausulerte statlige folketel-

lingene etter 1910 og de kommunale folketellingene.

HBR åpner muligheter for å transkribere enkeltposter og via en kildetabell kunne lenke postene både til personsider og til det skannede bildet av kilden. Senere brukere vil kunne sjekke kvaliteten ved å følge lenkene til skannet bilde. Dette vil øke tilgangen til transkriberte kilder og bedre kvaliteten i HBR, men det er samtidig en fare for at dette kan redusere interessen for fulltranskribering. Dersom færre frivillige påtar seg jobben med å transkribere hele kilder, vil vi tape viktige bidragsytere. En transkribering der brukeren ut ifra opparbeidet kunnskap har kjennskap til navnet på den som opptrer i kilden, åpner for bedre kvalitet på transkriberingen. Men det åpner også for feilkilder ved at man ikke legger vekt på mindre avvik.

## Oppfølging av brukere

HBR ligger åpent tilgjengelig på Internett. For å kunne bidra, må brukere registrere seg med fullt navn og rett identitet sjekkes via e-post. Alle endringer registreres med tid og fullt navn på den som utfører endringen. En bruker kan etablere varsel slik at vedkommende får en e-post om det skjer endringer på sider der brukeren har bedt om å bli varslet. Det er også mulig å sperre kontoer til useriøse brukere. Det er en stor bruker- og utviklergruppe som videreutvikler MediaWiki-plattformen som brukes på et stort antall nettsted over hele verden. Det er nødvendig med betydelig IT-kompetanse og erfaring med sosiale medier på Internett for å få et slikt nettsted til å fungere godt. Vi må være forberedt på å tilpasse utviklingen etter hvert som vi får erfaring fra brukeren. Det vil være en wiki med vesentlig flere sider enn engelsk wikipedia og med et stort antall brukere.

Innen Mediawiki er det lagt til rette for mange forskjellige måter å følge opp brukere

på, og wikipedia bruker dette til å lage et leksikon av høy kvalitet. For eksempel kan man etablere en gruppe superbrukere som får varsel når nye brukere gjør sine første endringer. På den måten kan man gripe inn tidlig med veiledning. Det er også mulig å låse sider, for eksempel ved stor uenighet mellom brukere slik at de ikke kan endres av andre enn de med administrative rettigheter. HBR vil forsøke å etablere superbrukere som følger opp forskjellige deler av nettstedet. Det mest naturlige vil være å etablere kontakt med lokale historielag som bidrar til oppbyggingen og følger opp endringer innen sitt område.

HBR ønsker å stimulere flest mulig brukere og legge til rette for at HBR brukes på forskjellige måter og med ulike innfallsvinkler. Brukere kan derfor etablere prosjekter som rendyrker arbeidet innen et tema. Prosjekter kan eksempelvis være drevet av et lokalhistorielag eller enkeltpersoner som fokuserer på en slekt eller personer med bestemte yrker. Prosjektsiden kan lenke til kilder eller prosjektsider utenfor HBR, og lenke til sentrale sider i HBR som prosjektet omtaler. Vi mener dette vil øke interessen for HBR, formidle informasjon om relevante prosjekter og gi honnør til bidragsyterne. Det primære språket i HBR er norsk, men de viktigste sidene vil også være tilgjengelig på engelsk. Det er betydelig interesse i utlandet både blant genealoger og andre forskere for å følge emigranter eller ta opp andre problemstillinger med utgangspunkt i en bedre organisering av de gode, norske kildene.

Det er ikke meningen at HBR skal erstatte annet arbeid eller dokumentasjon. HBR skal knytte sammen forskjellige typer informasjon og stimulere til å bygge videre på andres arbeid. Dette gjelder spesielt forholdet til Lokalhistoriewiki.<sup>39</sup> Lengre tekster og biografier skal fortrinnsvis legges til Lokalhistoriewiki eller andre eksterne kanaler enn HBR. HBR skal først og

fremst lenke informasjon, henviser til kilder og gi kortere livsløpsbeskrivelser.

HBR-wiki følger samme rettigheter som er vanlig i Wikipedia.<sup>40</sup> Det betyr at innhold kan kopieres, modifiseres og redistribueres så lenge brukerne gir *sine* brukere de samme rettighetene, og så lenge HBR krediteres som kilde. Men det legges begrensninger på bruk av bilder. De kan kun gjenbrukes om den som legger dem ut, gir tillatelse til dette.

## Enkelte øvelser ved Østerdalen

I Rendalsdatabasen kan vi få en pekepinn på hvilke utfordringer vi står overfor og hva slags resultater som kan forventes på lokalt plan. Av alle personer som er observert i perioden 1815–1824, er halvparten identifisert i dåp, vielse og begravelse, mens ytterligere 20 % er identifisert i dåp og begravelse, men ikke i vielse. På bakgrunn av den høye mortaliteten i perioden, må vi forvente at en stor del av de sistnevnte døde som ugifte. Rendalen var et prestegjeld med liten migrasjon og databasen er resultat av et grundig arbeid på et lite geografisk område. Vi kan ikke vente like stor dekningsgrad i kommuner med større migrasjon.

Selv etter at vi har lenket sammen flere personforekomster for individene, slik det er gjort i Rendalen, står vi overfor utfordringer når vi skal rekonstituere befolkninga på et gitt tidspunkt. Disse utfordringene blir ikke mindre av at mortaliteten var høy og at deler av befolkninga var geografisk mobil i de urolige tidene tidlig på 1800-tallet. Dødeligheten medførte at det ble inngått hele hundre ekteskap der enker eller enkemenn var involvert i Rendalen fra 1801 til 1815. Det er ikke noe enkelt puslespill å avgjøre hvor disse bosatte seg og hvilke barn fra tidligere ekteskap som fulgte med på flyttelasset. I Rendalen kan vi finne ut av dette ved å utnytte informasjon fra kilder som dekker seinere tidsperioder, for eksempel bosted ved dåp etter

1815. Slikt kildemateriale vil bare i begrenset grad være tilgjengelig i transkribert format for DNF1814. Da blir tilgang til lokal- og slekts-historisk ekspertise gjennom HBR gull verd. Det samme gjelder enkeltindivider med vanlige navn. Hvis vi ikke klarer å lenke Ole Olsen, 15 år i 1801-tellinga til den Ole Olsen som giftet seg i naboprestegjeldet i 1814, vil vi telle den samme personen flere ganger. Dette kan medføre en overestimering av folketallet dersom vi ikke i tilstrekkelig grad klarer å ta hensyn til manglende lenking i tellingen.

Det er vanskelig å si hvilken dekning vi kan oppnå i HBR. Det er også avhengig av i hvor stor grad man inkluderer «sannsynlige» lenker og ikke bare «sikre» lenker. I områder og perioder med stor migrasjon vil lenkingsgraden være lavere. Ved å ha en nasjonal database vil vi klare å fange opp deler av migrasjonen, men alt vil ikke være mulig å dokumentere i tilgjengelige kilder. Det må opprettes noen mål for lenkingsgrad som kan brukes til å sammenligne geografiske områder og perioder. Det kan være andel lenker mellom dåp og begravelse, andelen med en eller to foreldre identifisert, eller andel av personer i en folketelling som er lenket til minst én annen kilde.

## Oppsummering

HBR vil utgjøre en unik kilde til denne historiske perioden og vil kunne brukes i svært mange forskjellige forskningsprosjekter innen en rekke fag. I et internasjonalt perspektiv finnes det ingen sammenlignbare historiske befolkningsregistre med tilnærmet samme størrelse bygget ved å lenke flere ulike kildetyper. På samme måte som norske biobanker tiltrekker seg internasjonal forskning, kan HBR trekke internasjonal forskning til Norge for å studere norske data. Vi kan trekke på at Norge sammen med de andre nordiske landene har et godt tilfang av kildemateriale. En utstrakt

bruk av dette materialet krever utvikling av ny lenkingsmetodikk som består av en sammen-setning av flere etablerte teknikker kombinert med nye metoder som øker oversikten. Især er bruk av mediawikier en betydelig nyvinning på dette området. Det vil kreve stor dugnadsinn-sats fra et stort antall personer for å nå disse målene, men til gjengjeld gir dette oss en infra-struktur med stor bruksverdi for mange ulike forskningsoppgaver.

### *Lars Holden*

Adm. dir.

Adr.: Norsk Regnesentral  
Postboks 114 Blindern  
0314 Oslo  
E-post: lars.holden@nr.no

### *Gunnar Thorvaldsen*

Professor

Adr.: Registreringssentralen for historiske data  
Universitetet i Troms  
9037 Tromsø  
E-post: gunnar.thorvaldsen@uit.no

### *Torkel Rønold Bråthen*

Seniorrådgiver

Adr.: Arkivverket  
Postboks 4013 Ullevål Stadion  
0806 Oslo  
E-post: torbra@arkivverket.no

## Takk

Forfatterne takker Svetlana Boudko og Till Halbach Røssvoll ved Norsk Regnesentral som har implementert den første versjonen av HBR.

## Noter

- 1 Norsk Regnesentral. Lars.Holden@nr.no
- 2 Registreringssentralen for Historiske data, Universitetet i Tromsø. Gunnar.Thorvaldsen@uit.no
- 3 Arkivverket, prosjektleder for «Det norske folk i

- 1814». torbra@arkivverket.no
- 4 Bråthen 2011.
- 5 Første versjon av HBR er tilgjengelig på: [www.slekt.no](http://www.slekt.no), mens bakgrunnen for prosjektet beskrives på: <http://www.rhd.uit.no/nhdc/hpr.html>.
- 6 Thorvaldsen 2008 og Thorvaldsen 2011.
- 7 [http://no.wikipedia.org/wiki/Det\\_sentrale\\_folkeregister](http://no.wikipedia.org/wiki/Det_sentrale_folkeregister)
- 8 Thorvaldsen 2011.
- 9 NAPP prosjektet beskrives på: <http://www.nappdata.org/napp/>
- 10 Se <http://genetikportal.no/> for en beskrivelse av omfanget av dette arbeidet i Norge.
- 11 Hovig 2010.
- 12 Suren et al 2007.
- 13 <http://www.ddb.umu.se/om-ddb/publikationer/>
- 14 <http://www.pop.umn.edu/og> <http://www.nappdata.org/napp/>
- 15 Niemi et al. 2003: bind II.
- 16 Drake 1969: 8.
- 17 For informasjon om Mediawiki: <http://en.wikipedia.org/wiki/MediaWiki> og <http://www.mediawiki.org/wiki/MediaWiki>.
- 18 En personpost eller en personforekomst er en innførsel om en person i én kilde, for eksempel data om et barn i en folketelling eller om en mor i en dåp.
- 19 Professor Ólöf Gararsdóttir, Hagskola Islands, personlig kommunikasjon, november 2011. Se også: [www.decode.com](http://www.decode.com) og [http://en.wikipedia.org/wiki/DeCODE\\_Genetics](http://en.wikipedia.org/wiki/DeCODE_Genetics)
- 20 For Demografiska databasen, se: <http://www.ddb.umu.se>. For *Historical Sample of the Netherlands*, se: [www.iisg.nl/hsn/Bufret](http://www.iisg.nl/hsn/Bufret). For nettverket *European Historical Population Samples Network*, EHPS-net, se: [www.esf.org/ehps-net](http://www.esf.org/ehps-net).
- 21 Thorvaldsen 2008.
- 22 NAPP prosjektet beskrives på: <http://www.nappdata.org/napp/>
- 23 Thorvaldsen 2008.
- 24 [www.werelate.com](http://www.werelate.com) og <http://no.rodovid.org/wk/Hovedside>
- 25 <http://www.geni.com/>
- 26 Det følger av metoden for å generere databasen. WeRelate ser ut til å ha betydelig aktivitet for å redusere dette. Se blant annet diskusjonsforumet på: [http://www.werelate.org/wiki/WeRelate:Suggestions/Tweak\\_to\\_Duplicates\\_Report](http://www.werelate.org/wiki/WeRelate:Suggestions/Tweak_to_Duplicates_Report)
- 27 Hans P. Hosar, Lokalhistorisk institutt, personlig kommunikasjon.
- 28 Johansen 2002.
- 29 Variabelen «famsize» – antall medlemmer av egen familie i husholdet – kan analyseres sammen med andre konstruerte variabler hos Minnesota Population Center på nettstedet [www.nappdata.org](http://www.nappdata.org).
- 30 Programmene er utviklet i PL/SQ\_L et effektivt grensesnitt overfor RHDs databaser i Oracle. Navn sammenlignes ved hjelp av tekstalgoritmen Jaro-Winkler.
- 31 Thorvaldsen 1995.
- 32 Heimen 1/2005, spesialnummer om personvern og Thorvaldsen 2008.
- 33 1950-matrikkelen: <http://www.dokpro.uio.no/cgi-bin/stad/matr50>, 1886-matrikkelen: <http://www.rhd.uit.no/matrikkel/excel.html>. Matrikkelen fra 1838 er skannet, og vil forhåpentligvis snart være tilgjengelig som maskinlesbar og søkbar tekstversjon etter behandling med OCR.
- 34 HAG-prosjektet er et samarbeid mellom blant annet Kartverket, Arkivverket, Lokalhistorisk institutt og Institutt for folkehelse.
- 35 <http://bergis.uib.no/index.shtml>.
- 36 Solli 2006.
- 37 [www.werelate.com](http://www.werelate.com) har mye grafikk som blant annet benytter geografiske koordinater.
- 38 Eikvil et. al 2010 på: [http://www.rhd.uit.no/nhdc/HBR\\_notat\\_okt-2010.pdf](http://www.rhd.uit.no/nhdc/HBR_notat_okt-2010.pdf)
- 39 [www.lokalhistoriewiki.no](http://www.lokalhistoriewiki.no) er et nettsted drevet av Norsk lokalhistorisk institutt.
- 40 Opphavsrett på wikipedia beskrives på <http://no.wikipedia.org/wiki/Wikipedia:Opphavsrett>.

## Litteraturliste

- Bråthen, Torkel Rønold. «Det norske folk i 1814» i *Slekt og Data*, nr 4/2011, s. 44–45.
- Bull, Hans Henrik. *Data and Methods*. Universitetet i Oslo, 2006, s. 25–34.
- Datatilsynet, Slektsgransking, <http://www.datatilsynet.no/Teknologi/Internett/Slektsgransking/>
- Drake, M. *Population and society in Norway 1735–1865*. Cambridge, Cambridge University Press. 1969. Eikvil, Line, Holden, Lars og Bævre, Kåre. *Automatiske metoder som hjelp til transkribering av historiske kilder*. Notat SAMBA/44/10, Norsk

- Regnesentral. Oktober 2010.
- Gjelseth, M. *Relasjonsdatabaser som verktøy i en historisk-demografisk studie*. Hovedoppgave, Historiskinstitutt, Universitetet i Oslo, 2000.
- Heimen*, spesialnummer om personvern, 2/2005.
- Hovig, Eivind. «Norske gener?» i *Genialt*, nr 1/2010, s. 14–15. [http://www.bion.no/filarkiv/2010/07/Genialt\\_1\\_2010\\_norske\\_gener.pdf](http://www.bion.no/filarkiv/2010/07/Genialt_1_2010_norske_gener.pdf)
- Johansen, Hans Chr. «Identifying people in the Danish Past» I Sølvi Sogner, Hilde Sandvik, Kari Telste, Gunnar Thorvaldsen (red): *Pathways of the past: essays in honour of Sølvi Sogner on her 70th anniversary 15. March 2002*. Novus, Oslo 2002, s. 103–110.
- Niemi, Einar, Myhre, Jan Eivind og Kjeldstadli, Knut. «I nasjonalstatens tid 1814–1940» i *Norsk innvandringshistorie*, bind II. Oslo, Pax 2003.
- Arne Solli. *Urban space and household forms*. Artikkel presentert på the Eighth International Conference on Urban History, Urban Europe in Comparative Perspective, Stockholm 30th August – 2nd September 2006.
- Suren, Pål, Grijbovski, Andrej og Stoltenberg, Camilla. *Inngifte i Norge, Omfang og medisinske konsekvenser*, Folkehelseinstituttet, Rapport 2007:2.
- Thorvaldsen, Gunnar. «Using NAPP Census Data to Construct the Historical Population Register for Norway» i *Historical Methods*. Volum 44 (1) 2011, s. 37–47.
- Thorvaldsen, Gunnar. «Fra folketelling og kirkebøker til norsk befolkningsregister» i *Heimen*, bind 45, 2008, s. 341–359.
- Thorvaldsen, Gunnar. *Migrasjon i Troms studert i folketellingene 1866–1900*, Universitetet i Tromsø 1995.
- With, Christian 2005: «Slektsforskning og personvern». *Slekt og Data* 1, s. 11–13.