

Til: Arbeidsgruppen i HBR-wiki prosjektet

Fra: Lars Holden, NR

Dato 26.5.2010

## Status HBR-wiki

Dette er et arbeidsnotat og vil løpende bli oppdatert under utviklingen av prosjektet.

### Innhold:

Status HBR-wiki .....	1
Arbeidsgruppe .....	1
Langsiktig målsetning .....	1
Kvalitetssikring .....	2
Forhold til annen aktivitet innen feltet .....	2
Avgrensinger i data .....	3
Drift av og adgang til databasen .....	4
Ansvaret for og rettigheter til databasen .....	5
IT-plattform .....	5
Ambisjon forprosjekt .....	6
Transkribering .....	6
IDer .....	6
Eiendomskilde ID (EKID) og eiendomsid (EID) .....	8
Lenking og kobling .....	10
Grensesnitt .....	10
Sideoversikt i HBR-wiki .....	11
Utforming av person-, familie-, stedssider og kildetabeller .....	12

### Arbeidsgruppe

Lars Holden (leder, NR), Gunnar Thorvaldsen (RHD), Lars Nygaard (Arkivverket), Marianne Wiig (NLI), Anfinn Kjelland (HiVolda, BSS), Espen Berg (DIS-Norge) og Per Steinar Håklev (DIS-Norge). Det er åpent for å utvide med flere. Foreløpig planlegges det med at mye skjer pr epost heller enn å samle alle til møter.

### Langsiktig målsetning

HBR-wiki lenker sammen all åpen historisk person- og stedsinformasjon knyttet til personer i Norge.

Det er vanskelig å si hvilken dekning vi kan få til. Noen lokale databaser klarer opp til 90%. I områder og perioder med stor migrasjon, vil tallet bli 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 dagens kilder. Hvis det skal etableres kvantifisert målsetning, må man også formulere det som et presist mål. For eksempel andel av personer i en folketelling som er lenket til minst en annen kilde eller der man har identifisert en/begge foreldre.

## Kvalitetssikring

Kvaliteten sikres ved at å lage et mest mulig transparent system. Der det er mulig dokumenteres lenker med toveis pekere til kilde og HBR bidrar til ny transkribering der man lett ser både transkribert tekst og scannet bilde. Vi skal forsøke å etablere gode standarder og lokale grupper som mobiliserer til innsats og kvalitetssikrer. Alle endringer dokumenteres med person som utførte endringen og dato. Med wiki-teknologi er det lett å identifisere og reversere alle endringer utført av en person. Det er diskusjonssider for alle personer, familie/husholdninger og steder. Det kan være en like stor utfordring å kvalitetssikre de endringsforslagene som kommer, som å skape en respons. Vi regner med at det blir en meget besøkt database og mange som vil bidra til kvalitetssikring. Men som et dugnadsprosjekt vil det ta tid å få god dekning og høy kvalitet. Hvor mange som får være med i arbeidet og hvordan disse velges ut er ikke avgjort. Vi starter med en mindre gruppe for å vinne erfaring og vil så utvide gradvis og håper å kunne ha svært mange bidragsytere etter hvert.

## Forhold til annen aktivitet innen feltet

HBR er et forprosjekt finansiert av forskningsrådet motivert fra historisk, samfunns- og medisinsk forskning. Vi håper å få finansiert et hovedprosjekt fra forskningsrådet og etablere en permanent drift i samarbeid mellom flere institusjoner og foreninger. HBR vil være et non-profit aktivitet.

Målsetningen er at HBR-wiki skal bli et viktig verktøy for alle som jobber med historiske person- og stedsdata og stimulere til økt interesse for dette området. HBR-wiki skal lenke sammen data fra andre kilder og databaser og ikke erstatte disse. Det må vi ha et bevisst forhold til og etablere gode samarbeidsforhold med andre aktører. Istedenfor at stadig nye personer skal gjenoppdage de samme lenkene, så vil interessen gå i retning av å ha bedre dokumentasjon og kvalitet i lenkene, samle mer informasjon, gå enda lenger tilbake i tid og til enda flere kilder som emigrasjonsprotokoller, skattelister, konfirmasjonslister etc. som ofte inneholder mer informasjon enn en folketelling eller dåp. Konkret kan vi si følgende om forholdene til annen aktivitet og databaser:

1. Det skal være pekere til alle primærkilder, helst toveis elektronisk pekere, bl.a. Digitalarkivet. Der primærkilden bare er skannet, men ikke transkribert, kan det legges peker til skannet bilde og fortrinnsvis også til side der man kan bidra til transkribering av enkeltposter i kilden. (jf. Notat fra Kåre Bævre) HBR-wiki skal brukes aktivt for å øke interessen for transkribering både for åpne data og for data som bare vil være tilgjengelig i den lukkede delen.
2. Det vil være sider for hver person, familie/husholdning og bruk/plass/gård. Hver side vil ha mulighet for tekst, men det oppfordres til at fyldige tekster legges i Lokalhistoriewiki.no med toveis elektroniske pekere mellom wikiene. HBR har fokuset på referanser til kilder, men det er også mulig å legge inn tekster. Tekster i HBR er mer beskrivelse av livsløp og størrelse på gården. I Lokalhistoriewiki.no fokuseres det på tekster av mer allmenn interesse og de har bygget opp en stab som kvalitetssikrer dette arbeidet.
3. Hver person omtalt i en bygdebok bør ha en referanse til boken med henvisning til side eller helst elektronisk til skannet side. Referansen til bygdeboken bør fortrinnsvis være via Bibsys da dette samarbeidet dekker både Nasjonalbibliotekets og de fleste andre forskningsbiblioteks samlinger. Dermed får langt flere brukere lett adgang til eksemplarer.

HBR-wiki kan påvirke etterspørselen etter bygdebøker i begge retninger. Mange lesere vil være interessert i bygdebøker i bokform og få oppdatert informasjon etter 1910, generell omtale av bygden og annen informasjon i tekst og bilder som ikke er i HBR-wiki. Grensesnittet mot Bibsys kan være som i Lokalhistoriewiki.no.

4. HBR vil inneholde en lukket database for informasjon etter 1910. Når det skal utføres analyser på hele databasen, skjer det ved at HBR-wiki kopieres inn i den lukkede databasen. Da dette skjer på en bestemt dato, vil man ha bedre kontroll med hva som ligger i databasen.
5. Det skal ikke lagres informasjon om arvelige sykdommer i HBR-wiki. Her tenkes det ikke på dødsårsak som ligger i åpne kilder, men informasjon som ligger i lukkede databaser og som oppfattes sensitivt for nålevende personer. For noen typer forskning vil man lage en egen database om arvelig sykdommer og annen informasjon som ikke er offentlig tilgjengelig, men som bruker slektsforholdene og unike personIDer fra både åpen og lukket HBR.
6. Sidene i HBR-wiki vil indekseres og WeRelate ønsker muligheten til å lenke sine personer mot personene i HBR-wiki. Tilsvarende vil HBR-wiki ha muligheten til å lenke til personer i WeRelate, noe som vil gjøre det mulig å følge emigranter til USA.
7. HBR-wiki vil påvirke aktiviteten i DIS-Norge og annen frivillig slektsforskning. Vi antar at det vil øke aktiviteten og bedre kvaliteten i slektsforskning generelt. HBR-wiki åpner for samarbeid mellom mange mennesker og bedre dokumentasjon av resultater enn i dag. Med mange kilder og både fokus på personer og steder, vil det være tilnærmet ubegrenset av informasjon som skal lenkes. HBR-wiki trenger mange frivillige i lenkingen og personer som kan kvalitetssikre. Teknisk bistand er også ønsket.
8. WeRelate har en ambisjon om å lage grensesnitt til private wikier slik at disse kan utveksle informasjon. Det kan være aktuelt for HBR-wiki på sikt, men neppe noe HBR selv vil bruke ressurser på selv om det vil være av interesse for mange.
9. Det arbeides med å etablere en "historisk matrikkel", HAG-databasen for alle eiendommer fra 1600-tallet til i dag med informasjon om utskilling og sammenslåing av eiendommer. Vi planlegger å bruke informasjon herfra, men grensesnittet mot denne databasen er ikke avklart.
10. Mormonerne har et "Norway Project" med transkribering av data fra gårds- og slekts historier. De kan også brukes til kilde. Vi bør forsøke å få til en arbeidsdeling i transkriberingen.
11. Vi ønsker mest mulig åpenhet om prosjektet, men vi må i forprosjektet og under hovedprosjektet hele tiden understreke at det er under utvikling. Det innebærer at systemet vil ha mange svakheter og feil og vi har begrenset med ressurser til å ta imot bidrag.

## **Avgrensinger i data**

Målet er å registrere i HBR-wiki alle personer som har levd i Norge helt eller delvis i perioden 1800-1910 med referanser til kilder for denne perioden. Vi ønsker også referanser til innførsler om disse personene i kilder *før* denne perioden og i *åpne* kilder *etter* denne perioden. Vi ønsker også aner til denne gruppen som kan dokumenteres i sikre kilder, men har ikke samme ambisjon om å dekke alle personene. Her vil lenking i stor grad være synonymt med genealogisk forskning og bruk av Gedcom-filer fra seriøse genealoger *som eksempler* fra denne perioden vil være nyttig.

Etterkommere kan også inkluderes, men vi må være meget forsiktig med personvernet og vil håndheve 1910 grensen relativt sterkt. Samtidig vil lenking på 1900-tallet hjelpe betydelig for lenking av den lukkede delen.

### **Lovregulering:**

Statistikkloven setter 100 år grense for folketellinger. Forvaltningsloven setter en sperrefrist på kommunale folketellinger på 60 år ifølge Arkivverket. Kirkebøker har generell 60 års sperregrense. Der det er informasjon om adopsjon slik det kan være i dåps- og konfirmasjonslister, gir adopsjonsloven fra 1935 en sperrefrist på 100 år. Arkivverket har en 80 års grense på kirkebøker med gravlagte av hensyn til dødsårsak. Personopplysningsloven gjelder bare for levende personer, men setter også begrensninger for arvelige sykdommer. Det kan hevdes at lenking av offentlig kjent informasjon er å tilføre en merverdi som gjør at det ikke kan offentliggjøres.

Vi ser følgende alternativer:

1. Bare tillate informasjon frem til og med 1910 folketellingen. Ikke inkludere noen informasjon om arvelige sykdommer.
2. Utvide alternativ 1 med lenking til åpne lister over døde. (for eksempel <http://digitalarkivet.no/cgiwin/webcens.exe?slag=visbase&sidenr=10&filnamn=DODSFALL&gardpostnr=6#nedre>. Det bør ikke kreve godkjenning fra Datatilsynet og vil hjelpe betydelig i lenkingen på 1900-tallet.
3. Utvide alternativ 2 til løpende 80 års grense for åpne kilder, men begrense det til personer som er døde. Antar at personer er døde 100 år etter fødsel eller 80 år etter etablert familie. Det kan åpne for kommunale folketellinger, kirkebøker, "folkemengdens bevegelser", registerkort etc og redusere perioden for lukket HBR med 20 år. Men det kan hevdes at lenking av åpen informasjon er noe nytt slik at det vil kreve en godkjenning fra Datatilsynet.
4. Som 3 med 80 års løpende grense for åpne kilder, men ikke begrense det til personer som er antatt døde.
5. Som 4, men i tillegg søke om godkjenning for offentliggjøring av personinformasjon fra lukkede kilder (for eksempel folketellingen 1920 - 1950) for personer man vet ikke lever.
6. Kjente personer kan inkluderes for å øke interessen for databasen. Styret i HBR må lage en definisjon av kjente personer, for eksempel personer som omtales i leksikon eller wikipedia. Se [http://www.werelate.org/wiki/Category:Notable\\_people](http://www.werelate.org/wiki/Category:Notable_people). Lenking til Lokalhistoriewiki.no kan også brukes her.

### **Drift av og adgang til databasen**

Databasen etableres på UiT. Vi regner med at den flyttes til Digitalarkivet i hovedprosjektet. Lenker må gjøres dynamiske slik at det er lett å flytte sted. I forprosjektet er NR i samarbeid med UiT ansvarlig for det tekniske. Det tar lengre tid enn planlagt å installere programvaren på UiT pga problemer i koden.

I et hovedprosjekt må det etableres en driftsorganisasjon med historiefaglig og datakompetanse på flere årsverk. RHD/UiT kan etablere en foreløpig driftsorganisasjon inntil hovedprosjektet er

etablert. Når HBR-wiki er åpent tilgjengelig, vil det bli et betydelig behov for en driftsorganisasjon. Det vil være en fordel å etablere superbrukere for hvert geografisk område i samarbeid med lokale grupper.

Når database er etablert på slutten av forprosjektet, vil den ligge åpent tilgjengelig. Det vil være skriveaksess via login for relativt mange personer. I etableringsfasen vil skriveaksess utvides gradvis. Vanlige brukere vil typisk ha adgang til lenking og skrive biografier. Superbrukere vil følge opp endringer fra andre brukere. Det må være strenge regler for å endre data knyttet til kildene. Brukere identifiseres med fullt navn og inndeles i grupper med forskjellige rettigheter. Det vil være flere nivåer med rettigheter og endringer må forsøkes kontrollert på et høyere nivå. Brukere som misbruker rettigheter utestenges. Wiki-teknologien medfører at det er lett å se hvem som har skrevet hva og det er lett å reversere en endring. Nye brukere må følges opp spesielt.

### **Ansvar for og rettigheter til databasen**

Det vil bli flere prinsippsspørsmål under utvikling og drift av databasen. Noen aktuelle spørsmål er hva som skal registreres i databasen, hvem som skal gjøre det og med hvilke rettigheter. I et hovedprosjekt må det etableres et styre som har ansvaret for slike beslutninger. Inntil det er etablert, kan HBR-wiki arbeidsgruppen ta daglige avgjørelser, mens viktige spørsmål tas opp i HBR arbeidsgruppen og styret i forprosjektet.

Vi regner med at det blir mange bidragsyttere og stor interesse for databasen. De som er interessert i databasen kan grupperes i følgende grupper: forskere innen historie, samfunnsfag og medisin, slektsforskere og kommersielle interesser antagelig hovedsakelig innen genealogi (forfattere av bygdebøker, ancestry.com) og medisin. Dette krever at vi har et bevisst forhold til rettighetene. Vi bygger noe ikke-trivielt på toppen av andre kilder, så det vil være mulig å beskytte det vi skaper. Det er naturlig å kreve at bidragsyttere som bidrar til HBR gir sine bidrag under forutsetning av å følge HBRs rettighetsregler.

Alternativer rettighetsregler:

1. Som wikipedia at alt er åpent, men det skal refereres. Tekster kan fritt brukes til kommersielle formål. <http://no.wikipedia.org/wiki/Wikipedia:Opphavsrett>
2. Den som registrerer bilder kan reservere seg mot gjenbruk av bildet som i WeRelate [http://werelate.org/wiki/WeRelate:Terms\\_of\\_Use](http://werelate.org/wiki/WeRelate:Terms_of_Use)
3. Kreve at kommersiell bruk krever godkjenning. Godkjenning gis mot betaling som kommer arbeidet med HBR til gode. Kommersiell bruk regnes som å bruke en større del av datamaterialet for eksempel til bygdebok eller medisinsk forskning.
4. Kreve betaling om brukere ønsker tilgang til data på bestemt format.

### **IT-plattform**

Vi bygger på softwaren WeRelate. ([www.werelate.org](http://www.werelate.org)). Det er oppnådd enighet med WeRelate om avtalen for bruk og videreutvikling av WeRelate i en ikke tidsbegrenset og lisensfri avtale. Både WeRelate og HBR-wiki vil ha gjensidig rett til å bruke utvidelser. Vi har mottatt koden 23.3 og UiT har startet installeringen. Det er noen problemer med konsistens i koden og installeringen har tatt en del tid (pr 6.5.2010)

Vi planlegger at HBR-wiki blir engelsk og norsk språklig. Viktigste sider og systemkommandoer blir på begge språk. Mange av sidene vi kopierer fra WeRelate blir bare på engelsk, mens nye sider og alle personsider antagelig bare blir på norsk. WeRelate er et stort og avansert dataprogram og består av 4 separate deler, PHP utvidelser av mediawiki, Java indeksering og søkemaskin bygget på SOLR, Java GEDCOM uploader og diverse mindre vedlikeholdsprogram i Java.

### Ambisjon forprosjekt

1. Etablere en første versjon av softwaren.
2. Lese inn mest mulig av den historiske matrikkelen og opprette stedssider.
3. Lese inn store folketellingene fra 1801-1910 og etablere person og lenke til stedssider  
Forsøke automatisk lenking til flest mulig steder ved hjelp av gnr/bnr.
4. Importere automatisk lenking og lenking fra noen lokale databaser
5. Åpne for at brukere kan lenke interaktivt i wiki-teknologien.

Dette er meget ambisiøst i forhold til ressursene i forprosjektet. Vi må se hva som er realistisk å få til. Det er viktig å være landsdekkende i forprosjektet fordi det vi gi et helt annet løft i lenkingen mens vi venter på hovedprosjektet. En ekstra utfordring er et permanente ID og URLer i Digitalarkivet og formatet på HAG-databasen ikke er avklart.

### Transkribering

Det er mange kilder som ikke er transkribert. Vi antar at HBR vil bli meget populært og kan bidra til å øke interessen for transkribering. For åpne kilder vil det være ønskelig å gi elektronisk tilgang til scannet bilde og tilhørende tekstboks der brukere kan skrive inn transkribert tekst for et og et felt av gangen. Senere bruker vil se både transkribert tekst og scannet bilde og kan verifisere transkriberingen. Så snart et felt er transkribert vil det kunne brukes i søking. Det vil også kunne legges inn lenker fra personforekomsten til person og stedssider i databasen for å kunne lenke personopplysningene med andre data.

Vi planlegger at HBR også kan bidra til transkribering av lukkede kilder. Brukere vil bare kunne se isolerte felt. Det kan motiveres til transkriberingen ved å offentliggjøre statistikk fra transkriberingen, etablere konkurranser og kanskje også betaling. Kvaliteten kan sikres ved å av å til be om transkribering der man kjenner svaret.

### IDer

**Terminologi** En *personforekomst* (benevnt person-kilde-ID PKID) er en post i en transkribert (e.a.) kilde der én person er nevnt med navn og/eller andre opplysninger. En *personenhet* (forkortet PID) er et dataobjekt tilhørende en identifisert enkeltperson, inneholdende en unik person-ID, oppsamlede, korrigerte opplysninger om personen, og med lenker til *forekomstene* av denne personen i kildene og til nære slektninger (mor, far, partnere, barn). Tilsvarende etableres eiendom-kilde-ID EKID og eiendom-ID, EID.

### Digitalarkivet gir følgende opplysninger om endringene i Digitalarkivet

*"Digitalarkivet vil etablere ny programvare innen utgangen av 2010. Vi vil trolig satse på standardiserte XML-filer som de offisielle versjonene av datafilene. Det er altså her korreksjoner*

*av transkriberingsfeil o.a. skal utføres, med påfølgende import av ny versjon i Digitalarkivets database. Denne databasen vil logisk bestå av handlingsposter fra kirkebøker og mange andre kildetyper, pluss personposter (én person per post). Fra folketellingene og noen andre kilder vil det istedenfor handlingsposter være bostedposter på flere nivåer; tellingskrets, gård og leilighet (i byene). Poster av alle typer vil få en unik ID. IDene vil inngå i en URN, som i sin tur vil inngå i en permanent weblenke (URL) til hver enkelt post. Det vil følgelig være mulig å oppgi permanente referanser til postene både fra andre nettsteder (f.eks. HBR) og i trykte publikasjoner. Fra de transkriberte dataene i Digitalarkivet vil det også være lenker til den skannede versjonen av kilden (om mulig til den aktuelle siden i kilden) og til beskrivelsen av kilden i Arkivportalen, som en felles nettingang til alle landets arkivkataloger basert på Asta 5.*

*IDene vil inngå i en URN, som i sin tur vil inngå i en permanent weblenke (URL) til (person)postene i Digitalarkivet. Vårt altoverskyggende fokus er å sikre at IDene blir unike, og at de ikke skal måtte endres senere fordi data som kan inneholde feil, er tatt med i IDen. En slik ID i Digitalarkivet vil f.eks. se slik ut: pf20101217000034, der pf står for personforekomst, 20101217 er datoen IDen ble opprettet, og 000034 er et løpenummer innenfor denne datoen.*

*URLene til personforekomster i dagens Digitalarkiv (inneholdende "webcens" osv.), kan ikke betraktes som permanente. De vil forvinne når det "nye Digitalarkivet" overtar alt materiale i 2011. De PKIDene vi vil tildele fra sommeren i år og fremover, vil imidlertid være permanente. URL til personforekomstene i det nye Digitalarkivet vil bli standardiserte med PKIDen som eneste variable ledd, men "layouten" på URLen er ikke klar ennå. "*

### **System for person-ID'er PID**

HBR planlegger å bruke en regelbasert definisjon av PID ut fra PKID. Vi lager en prioritetsregel for PKID. PID er definert direkte ut fra PKIDen med høyest prioritet av de som er lenket til den aktuelle personen. Dette er et system som lett håndterer endringer i lenking. Det kan illustreres ved at PKIDene A og B er lenket og så finner vi en tredje PKID C som vi lenker til B og som viser at A og B ikke kan være lenket. Da endrer vi bare PIDene fra A til A og B. Alle som referer til en person refererer til PKIDene A, B eller C eller en tilsvarende PID og HBR vil hele tiden vise videre til den PIDen med høyest prioritet av disse hvis man spør om PID. Det kan være at URL'en til personsiden i praksis vil være PIDen og prioriteringsreglen bestemmer hvordan sider slås sammen ved lenking.

Forslag til prioriteringsrekkefølge for PKID'er

1. folketelling 1910,
2. dødslistene før 1910,
3. fødsler etter 1910,
4. liste over utvandret før 1910
5. liste over innvandret etter 1910
6. andre folketellinger der senere folketellinger har høyere prioritet enn lavere.
7. andre oppføring i kirkebok der senere oppføring har høyere prioritet enn lavere
8. andre kilder der senere oppføring har høyere prioritet enn lavere.

De første fem kategoriene er ikke overlappende og tre første er antagelig av relativt god kvalitet. Folketellingen 1910 brukes til å lenke mellom HBR-wiki og lukket HBR.

Vi kan ikke regne med at det er overensstemmelse med kjernedata (navn, alder etc. ) mellom disse kildene. Der vi velger kjernedata automatisk, så må vi ha en annen prioriteringsregel basert på forventet kvalitet. Det er foreslått følgende prioritering:

a) kb dåp, b) kb konfirmasjon, c) kb vielse, d) kb begravelse, e) ft 1910, f) ft 1900, g) ft 1891 osv. Det er mulig d og e bør bytte plass. Det vil imidlertid være mulig å manuelt å skrive inn kjerneopplysninger.

## Eiendomskilde ID (EKID) og eiendomsid (EID)

HAG prosjektet, *Nasjonale karttjenester for historiske, administrative grenser* (HAG) nedsatt av bl.a. Kartverket avsluttes ved utløpet av 2010. Det lager en database som (i teorien) vil inneholde alle verdslige, geistlige, juridiske og administrative områder (på alle nivåer) som har eksistert i Norge i perioden mellom 1600-tallet og vår tid. Kåre, Lars N og Marianne deltar i dette prosjektet. I tillegg til startår, sluttår, endringshistorikk, navnehistorikk, kartfesting m.m. vil disse områdene også få sin unike ID i denne databasen. Det bygger på Kåres enorme arbeid med de historiske, administrative områdene. Det er en ambisjon om å utvide til en "historisk matrikkel" eller et historisk eiendomsregister for Norge.

Det nye Digitalarkivet vil opprette en (søke)tabell med "eiendomsforekomster" på samme måte som for personforekomstene. "Intetsigende" IDer vil bli tildelt disse på samme måte, mens det som måtte finnes av annen identifiserende informasjon som navn, matrikelnummer, løpenummer, gårdsnummer, bruksnummer osv. vil bli plassert i andre felt i tabellen. Bosteds- og eiendomsinformasjon fra følgende kilder vil *etter hvert* bli plassert i denne tabellen: matrikler 1660-1950, bosteder i folketellinger 1865-1910 (+ 1801?), bosteder i skifteregistre, gårder og bruk i panteregisterregistre, metadatabasen til grunnboksbladene 1936-93, GAB 1990, eiendomsregistret til jordskiftekartene, bosteder i manntall?, bosteder i skattelister? m.m. Oppgitte bosteder i bl.a. kirkebøkene og emigrantprotokollene vil trolig *ikke* bli lagt inn, siden denne informasjonen (kun gårdsnavn etc.) alene er for "tynn" for presis eiendomsidentifisering. Riksarkivet kan etter hvert tilby eksporter fra eiendomstabellen om gårder, bruk og andre eiendommer for import i HBR som grunnlag for lenking av eiendomsforekomster.

Gårdene er relativt stabile over tid og fra en wikiside for hver gård er det mulig å lage en oversikt over endringer av bruk under gården. Flytting mellom administrative enheter vil ikke skape særlig problemer ved at wiki-siden for gården er den samme og det er toveis pekere til de aktuelle administrative enhetene over. Når et bruk skilles ut etableres toveis pekere mellom nytt bruk og bruket det skilles ut fra. Vi kan lese inn 1886 og 1950 matrikkelen og knytte stedssidene fra folketellingene til disse stedene. I følge Gunnars bok fins gårdsnavn/gatenavn i alle folketellingene, gårdsnummer/ husnummer/ gatenummer i 1801, 1891, 1900, 1910 og matrikelnummer i 1865, 1875 og 1910 (bare byene). Kåre sier at 80 til 90 % av gårdene kan det knyttes entydige matrikelnummer til. Hvis vi klarer så mye i et forprosjekt vil det være meget bra og resten kan håndteres i et hovedprosjekt. Både restkategorien på ordinær landsbygd, Finnmark som ikke ble matrikulert før etter vår åpne periode, og byene vil kreve omfattende spesialarbeid, især sistnevnte pga overgangen fra løpenummer i byen til løpenummer i gata. Arne Solli, UiB, har gjort et imponerende arbeid med person og stedsinformasjon for byene. I et hovedprosjekt kan det bli en viktig input.

1886 matrikkel <http://www.rhd.uit.no/matrikkel/excel.html> har bl.a. kolonner for kommunenummer, etternavn, fornavn, gnr, bnr, mnr, lnr, gårdsnavn og bruksnavn. Denne kilden er spesielt viktig siden den knytter sammen den gamle nummereringen mnr/lnr og den nye nummereringen gnr/bnr. Alt dette kan legges inn i en kildetabell med peker tilbake til transkribert kilde og peker til personside for eier og stedsside for bruk. Siden for hvert bruk kan ha pekere hierarkisk til sider for gård, kommune og fylke. På kommunenivå kan det legges peker til denne delen av kildetabellen. Når vi leser inn folketellinger med gnr/bnr eller mnr/lnr, kan vi automatisk lenke til stedssidene (enten slå sammen sidene eller lage mulig dublettpeker.)

Matrikkelen for 1950 inneholder ca 85.000 eiendommer i Norge, <http://www.dokpro.uio.no/cgi-bin/stad/matr50>, men har ikke med Finnmark og bykommuner.

### **Beskrivelse av planer for eiendommer i HBR-wiki.**

Med utgangspunkt i HAG-databasen kan Kåre lage oversikt over alle gårds- og bruksnummer i Norge med oversikt over alle endringer, hovedsakelig oppsplitting i nye bruk, fra 1600-tallet og frem til i dag. Kåre kontakter Kartverket og finner ut om vi kan legge ut alle bruk per 1.1.2010 eller om vi skal ta utgangspunkt i en tidligere dato. Han vil også finne ut hvilke data som vi kan legge ut i tillegg til navn, gårdsnummer, bruksnummer, endringshistorikk og hierarkisk lenking til gård og den til enhver tid gjeldende inndeling i kommuner og fylker. Hvis vi fører oversikten frem til 2010, får vi svært mange bruk som ikke fantes i 1910. Men mange vil være interessert i å følge historien til dagens bruk. Vi håper også å kunne legge ut bl.a. gateadresse for de eiendommene som har dette og koordinater. I løpet av sommeren vil Kåre også lenke disse brukene med eiendommene i folketellingene med unntak av 1801-tellingen som vil ta lengre tid siden denne bruker gårdsnavn og ikke mnr/lnr. I HBR-wiki etablerer vi en kildetabell med alle brukene. Kildetabellen vil ha felt for adm. inndeling, gnr/bnr, mnr/løpenummer/litra, peker til HAG-databasen, peker til matrikkelside og når og fra hvor bruket ble skilt ut og når og fra hvor brukes evt. ble slått sammen med. Matrikkelside er en wikiside for hver av disse brukene med lenker til kildetabellen. Disse vil være faste og bare administrator kan endre disse. Men det vil bli lagt ytterligere informasjon på disse sidene når vi finner kilder som refererer til disse stedene. Mest aktuelt er folketellingene, men senere kan det også være matrikler (bl.a. 1886) og kirkebøker. Når vi leser inn folketellingene i HBR-wiki, vil vi enten lagre den nye informasjonen (peker til familieside, kildetabell og transkribert kilde) på matrikkelsiden eller vi vil opprette en ny stedsside og etablere mulig dublettpekere til matrikkelsidene der vi klarer å etablere disse. Der en folketelling har en annen administrativ inndeling enn vi har i dag, vil vi lage et hierarki med de administrative enhetene som fantes på det tidspunktet. Dette hierarkiet med administrative enheter vil integreres med de administrative sidene til matrikkelsidene som beskrevet i kapittelet for Sideoversikt i HBR-wiki, Stedsinndeling.

Det er sannsynlig at vi har en kildetabell før 1886 med mnr/lnr og en etter 1886 med gnr, bnr. For begge periodene har vi følgende forslag til felt:

Fylke, Kommune, gårdsnavn, bruksnavn, (alt mnr/lnr) gårdsnummer, bruksnummer, gateadresse, koordinater, peker stedsside, peker HAG databasen

Hvilket gård/bruk det er skilt ut fra:

År, fylke, kommune (på tidspunkt for utskilling), gårdsnummer, bruksnummer, (alt mnr/lnr), peker stedsside

Hvilket gård/bruk det er gått inn i:

År, fylke, kommune (på tidspunkt for sammenslåing), gårdsnummer, bruksnummer, (alt mnr/lnr), peker stedsside

Der det er endringer i administrativ inndeling på kommunenivå må det i mange tilfeller registreres som ny enhet.

## Lenking og kobling

I noen av beskrivelsene trenger vi et økt presisjonsnivå for det vi litt upresis kaller lenking. Da bruker vi pekere om referanser til andre websider, lenking om å føre sammen opplysninger om samme person i forskjellige kilder og kobling om å knytte personer sammen i kjernefamilier eller andre relasjoner. Vi vil bruke forskjellige metoder i arbeidet:

1. En første kobling vil følge direkte ved etablering av personsidene fra primærkildene. Personer i samme husholdning vil bli koblet ved etablering. Gunnar jobber med å identifisere roller i husholdningene og entydighet i stedene ut fra data i Minneapolis og RHD.
2. Miljøet i Minneapolis (Gunnar har kontakt) prøver ut automatisk lenking ved Febrl-programvaren og Kåre B. arbeider med tilsvarende programvare. Vi må finne en egnet måte å lese en slik lenking inn i HBR-wiki. Mulig dette kan lagres som Gedcom filer eller definere nytt format der hver linje har 2 PKID og så lages mulig dublettpeker mellom disse personsidene.
3. Lokale databaser fra BSS, Asker og andre steder. Noen av disse databasene vil være av høy kvalitet og vi ønsker å slå sammen personsider og forsøke å bevare alle PKIDene for hver person. Dette kan leses inn som Gedcom filer. Men det er usikkert om vi automatisk klarer å lenke med personsidene fra folketellingene eller hvordan kildeinformasjonen lagres i disse personsidene.
4. WeRelate har programmer for lenking. I dag foreslås lenking av alle personsider som er på hver brukers "watch list". Lenkingen må så gjøres manuelt i hvert enkelt tilfelle. Dette kan brukes i den manuelle lenkingen og koblingen. Det er mulig å lese inn Gedcom filer av høy kvalitet og med mange lenker og koblinger, men uten noen unike PKIDer. Disse kan brukes i en manuell lenking og kobling i HBR-wiki.

## Grensesnitt

Det må etableres presise grensesnitt mot andre databaser

1. Lenking til lukket HBR kan skje ved at begge databasene lenker til personene i 1910 folketellingen. I forprosjektet vil vi imidlertid ikke etablere programvare for å sette sammen disse databasene.
2. Grensesnitt til BSS og relasjonsdatabaser.
3. WeRelate har en mulighet til å legge inn Gedcom filer. Jeg tror ikke vi vil klare å koble innlesning av Gedcom til PKID/kildetabellen. Når man leser inn Gedcom filer vil man da få inn mange personer som ikke er lenket til PKIDer, men til gjengjeld lenket til mange andre personer. Når man lenker manuelt kan man kombinere denne informasjonen ved å søke i begge grupper av personer. Vi må stille høye kvalitetskrav til de Gedcom filene som leses inn slik at vi beholder god kvalitet også på personer uten PKIDer. Det vil være lett å gjenkjenne personer som er lenket til kildene og personer som kommer fra Gedcom.

4. Det er mulig vi bør etablere et meget enkelt format som gitt to PKIDer eller pekere til personsider, lenker eller kobler disse to personforekomstene. Det kan for eksempel brukes til å registrere forslag til endringer fra brukere som ikke har skriveaksess.

## Sideoversikt i HBR-wiki

<**Forsiden**> Kort beskrivelse av prosjektet. Lenker (i søyle på venstre side på forsiden): om HBR, kontakt, lenker til kilder, prosjekter, bidrag til HBR, søking, stedsinndeling, lenkesamling

<**Om HBR-wiki**> mer utfyllende om prosjektet.

<**Nyheter**> Kronologisk oversikt over nyheter. De siste nyhetene bør ligge på forsiden.

<**Kontakt**> Det bør etableres et nettforum der man kan diskutere HBR-wiki. Nettforum med temainndeling er en effektiv måte å håndtere mange diskusjoner parallelt og der tidligere svar bevares. WeRelate har et nettforum i dag som vi må tilpasse til vårt formål. Se også Digitalarkivets debattforum og [www.disnorge.no/slektsforum](http://www.disnorge.no/slektsforum). Det må i tillegg være mulig å henvende seg pr epost og være en ansvarlig organisasjon bak.

Det vil være fint om vi etter hvert klarer å få lokale DIS-Norge grupper til å håndtere noen spørsmål. Hvis personer som ikke har skriveaksess har konkrete forslag til oppdateringer, vil det være effektivt om vi finner et standardisert format for dette, for eksempel lenker til to personsider som skal lenkes. Da vil det være lett å legge inn mulig dublettpekere automatisk.

<**Peker til kilder**> Oversikt over kildetabeller der det er toveis pekere mellom kilder og person/stedssider. Henviser videre til sider for hver kilde, for eksempel hver folketelling, kirkebøker. For de viktigste kildene er fulle tabeller for hele kilden. Men for kilder der ikke alle skal registreres, for eksempel amerikanske immigrasjonslister, legges det inn linjer etter hvert som det er behov for det.

<**Prosjekter**> Oversikt over prosjekter. Prosjekter kan være å samle data for en region/sted (kommune, sogn, gård), familie, kjente personer eller andre temaer som etniske grupper og emigrasjon. Rene fagartikler som er av interesse for brukere av databasen kan også legges her, for eksempel mer utfyllende informasjon om hvordan man lenker, navnetradisjoner og resultater fra bruk av databasen i historisk forskning.

Vi må regne med mange prosjekter for områder, familier og steder. Det blir dermed et hierarki av sider knyttet til områder. Familiesider knyttes både til etternavn nasjonalt og stedssidene (kommune). Hvert prosjekt får en side som beskriver aktiviteten. Derfra er det igjen lenker til personer og steder i databasen.

Prosjektsider kan enten være åpne for å redigere for alle med skriveaksess, for eksempel råd om hvordan bidra i lenkingen, til bare åpne for prosjektdeltagerne for eksempel resultater fra et bestemt forskningsprosjekt.

<**Bidra til HBR**> Hvordan man kan bidra til HBR, dvs. lenking, kobling, skrive biografier, legge inn bilder eller lignende. Her vil mulighetene endre seg i løpet av prosjektet. I begynnelsen

ønsker vi kontakt med personer som kan gi store bidrag enten i form av lenkede databaser, softwareutvikling eller kvalitetssikre data innen en region.

<**Søking**> informasjon om hvordan det søkes i databasen

<**Stedsinndeling**> Oversikt over alle steder i databasen. På øverste nivå er det liste over alle fylker og lenke til side for utlandet. Vi vil ikke legge arbeid i steder utenfor Norge, men i forhold til emigrasjon/immigrasjon, vil det være en fordel også å ha wiki-sider også for steder i utlandet. På hver stedsside er lenker til områder/steder på et lavere nivå, dvs fylke, kommune, sogn, gård, bruk/plass. Inndelingen avhenger av tidsperiode fordi grenser flyttes. Vi må etter hvert ha en full oversikt over alle inndelinger med tidsperioder. Man kan komme til samme side for en gård ved å følge lenker via forskjellige kommune og sogn pga inndeling i forskjellige tidsperioder.

Dersom det ikke er for mange forskjellige inndelinger så kan oversikten være på formen

- område A (1820-1900)
- område B (-1760)
- område C (1760-)

Hvis det er store endringer er det mer oversiktelig med egne sider eller lister for inndeling i de forskjellige periodene.

Siden de aller fleste personopplysninger er knyttet til steder, vil det være mulig å finne personer ved å lete etter stedet. Når det gjelder utlandet må vi bygge opp et hierarki med sider etter hvert som det kommer flere stedsnavn.

<**Lenkesamling**> Oversikt over relevante nettsteder bl.a. der det er mange kilder og andre wikisteder.

## Utforming av person-, familie-, stedssider og kildetabeller

En **typisk personside** i WeRelate ser slik ut.

[http://www.werelate.org/wiki/Person:Hans\\_Myhre\\_%282%29](http://www.werelate.org/wiki/Person:Hans_Myhre_%282%29)

I HBR bør vi beholde omtrent samme formatet. Navnet på siden genereres ut fra personnavnet i kilden med høyest prioritet og så et løpenummer for dette navnet. Dette håndteres automatisk av WeRelate.

Til venstre på siden er det en søyle med kjerneopplysninger. Disse kan velges fritt og kan endres manuelt eller på en automatisk måte. De er ikke koblet til sidenavnet. Når vi genererer sidene automatisk, velger vi kjerneopplysninger ut fra prioritet for kildekvalitet. Sted beskrives som bruksnavn, gårdsnavn og kommune. Alle som bor i husholdningen har peker til familiesiden.

Ved sammenslåing og splitting av personsider bruker vi følgende regler:

Kjerneopplysninger tas alltid automatisk fra kilde for høyest kildekvalitet. Hvis man mener det er feil i denne, kan det rettes manuelt. Når sider slås sammen får den nye siden manuelt innlagt tekst fra begge sidene. Når sider splittes opp, beholder siden med høyest prioritet teksten som er

lagt inn, dvs. at den blir i den siden den er og flyttes ikke til den nye siden. Den må evt. flyttes manuelt.

Nederst på siden er det plass for tekst og bilder (biografi/livsløp). Fyldige biografier ønsker vi skal ligge i NLI wiki. Øverst på siden er det en liste for mulige dubletter og kilder. En mulig dublettliste kan være slik

Fornavn etternavn	Fødselsdato	Peker til personside	Begrunnelse for lenking
Fornavn etternavn	Fødselsdato	Peker til personside	Begrunnelse for lenking

Begrunnelse for lenkingen er et rent tekstfelt og det bør stå når og hvem som er ansvarlig (person/prosjekt/dataprogram og bruker) for lenkingen. Alle wikisider har en side der man kan diskutere siden. Det er også mulig å legge informasjon på denne diskusjonssiden. Lange begrunnelser som viser til en navnetradisjon, hvem som var faddere i en dåp etc. er bedre å legge på denne siden.

En kildeliste kan være som under (kilde, navn, fam. stilling, sivilstand, fødsel, peker til gård, familie, kildetabell og transkribert kilde ). Rekkefølgen er kronologisk.

Dåp	Fornavn etternavn	Dåpsbarn	ug	15. august 1860	Peker sted	Peker familie-side	Peker kildetabell	Peker transkribert kilde	Begr .
Folketelling 1865	Fornavn etternavn	Sønn	ug	1860	Peker sted	Peker familie-side	Peker kildetabell	Peker transkribert kilde	Begr .
Folketelling 1875	Fornavn etternavn	Tjenestegutt	ug	1859	Peker sted	Peker familie-side	Peker kildetabell	Peker transkribert kilde	Begr .
Folketelling 1910	Fornavn etternavn	Husfader	g	10. august 1860	Peker sted	Peker familie-side	Peker kildetabell	Peker transkribert kilde	

For- og etternavn bør være i egne kolonner for å lette søk. Dato er for fødsel. Peker kildetabell er til en bestemt linje i tabellen. Peker familieside brukes for alle som er knyttet til husholdningen. "Begr." er begrunnelse for lenking og viser fra kilde med lavere prioritet til kilde med høyere prioritet. Det fremgår imidlertid ikke nødvendigvis hvilken kilde med høyere prioritet det er lenket til. Det vil imidlertid fremgå av kildetabellen. Vi trenger ikke begrunnelse for folketellingen 1910 da denne har høyest prioritet. Det er de andre kildene som trenger begrunnelse for lenking til denne kilden.

En **typisk familieside** (eller mer dekkende husholdningsside) i WeRelate har formen

[http://www.werelate.org/wiki/Family:Erik\\_Romsaas\\_and\\_Eli\\_Seielstad\\_%281%29](http://www.werelate.org/wiki/Family:Erik_Romsaas_and_Eli_Seielstad_%281%29)

Til venstre er liste med kjerneopplysninger med lenker til alle familiemedlemmer. I den store ruten til høyre på familiesiden etableres tabell for alle andre personer tilknyttet husholdningen. Tabellen kan være på formen under (kilde, navn, fam. stilling, sivilstand, kjønn, fødsel, lenker til gård, familie, kildetabell og transkribert kilde ). Rekkefølgen er kronologisk.

Folketelling 1865	Fornavn etternavn	Hennes mor	ug	K	15. august 1800	Peker person	Peker kildetabell	Peker transkribert kilde
Folketelling 1910	Fornavn etternavn	Tjenestegutt	ug	M	10. august 1890	Peker person	Peker kildetabell	Peker transkribert kilde
Folketelling 1910	Fornavn etternavn	Tjenestejente	ug	K	1888	Peker person	Peker kildetabell	Peker transkribert kilde

Lenkingen følger av personsidene. To familiesider bør automatisk slås sammen hvis vi klarer å lenke både far og mor i de to familiene. Det opprettes også familiesider for barn født utenfor ekteskap og ved ekteskap der man bare kjenner den ene av foreldrene.

Det er flere grunner til å legge inn hele husholdningen i familiesiden: Lenken fra stedssiden går til familiesiden og mange i husholdningen er knyttet til familien som besteforeldre, søsken av foreldre og barn av tidligere parforhold. I tillegg gjør det den automatiske innlesningen fra folketellingen mer robust ved at familiemedlemmer som ikke fanges opp automatisk, legges til resten av husholdningen istedenfor å utelates.

Nederst på siden er det plass for tekst og bilder. Fyldige tekster ønsker vi skal ligge i NLI wiki.

En **typisk stedside** for en gård i WeRelate er slik

[http://www.werelate.org/wiki/Place:Holte%2C\\_Drangedal%2C\\_Telemark%2C\\_Norway](http://www.werelate.org/wiki/Place:Holte%2C_Drangedal%2C_Telemark%2C_Norway)

navnet på siden er gårdsnavn, sogn, kommune/prestegjeld, fylke/amt. Stedssidene organiseres hierarkisk. På kommune nivå legges lister til gårder ut fra gnr/bnr og mnr/lmr som hentes fra matrikkelen.

Til venstre på siden er det en søyle med kjerneopplysninger. Disse kan velges fritt og kan endres manuelt eller på en automatisk måte. De er ikke koblet til sidenavnet. Når vi genererer sidene automatisk, velger vi kjerneopplysninger ut fra kilden. Gårdssiden lenkes igjen til kommune og fylke (og evt. finere inndeling) ved generering. Inndelingen følger folketellingen, bl.a. ser vi ikke forskjell på de forskjellige plassene på gården.

I det store feltet til høyre på siden legges liste for mulige dubletter (av sted) og kilder. En mulig dublettliste kan være slik

Stedsnavn	Peker til stedsside	Begrunnelse for lenking
Stedsnavn	Peker til stedsside	Begrunnelse for lenking



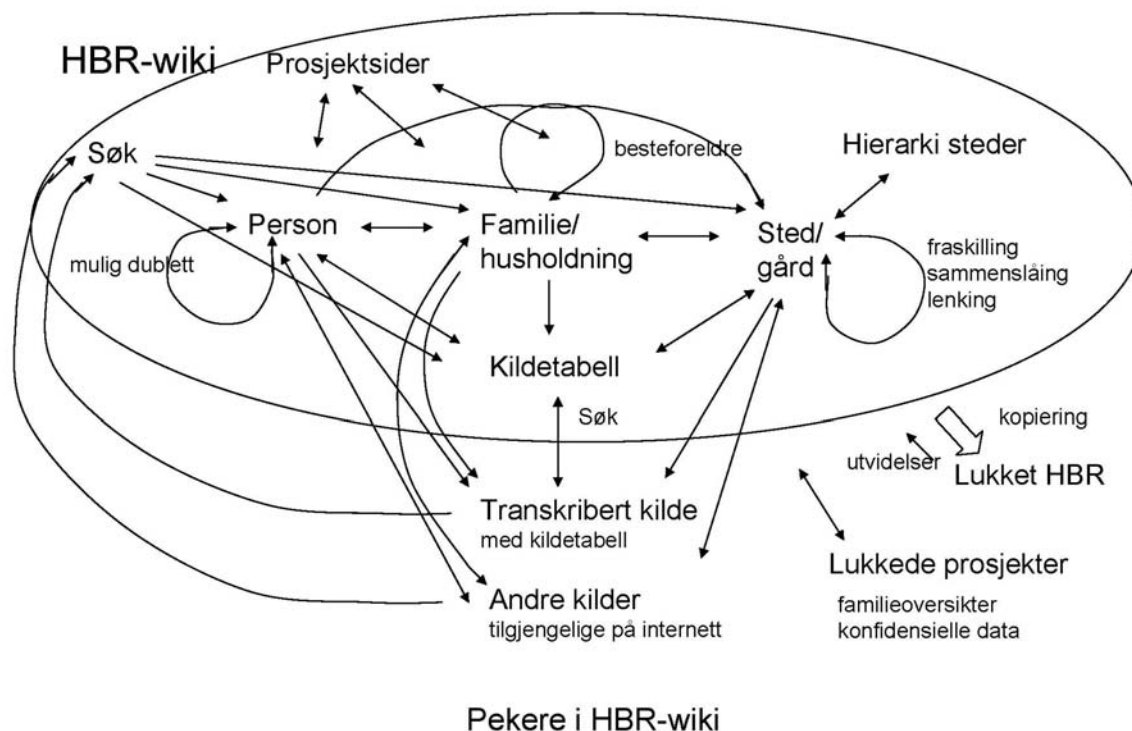
kolonner for geografisk inndeling, bla. for at løpenummer skal gi mening. Peger kildetabell er en peker til en personforekomst for samme person i en annen kildetabell med høyere prioritet. Den er derfor unødvendig i de kildene som er høyest prioritert. Når nye Dataarkivet etableres, må peker til transkribert kilde oppdateres og PKID etableres. Når en bruker er i for eksempel Digitalarkivet, kan man finne tilsvarende personforekomst her ved å søke på PKID eller et hvilken som helst annet felt.

Den eneste informasjonen som endres i denne tabellen ved sammenslåing av person- og stedssider er peker personside og peker kildetabell. Det er en fordel om dette gjøres automatisk og siden låses bortsett fra for superbrukere.

For kirkebøker må vi finne ut om vi ønsker en tabell pr kirkebok eller legge alle sammen inn i samme tabell. Ved en dåp vil alle personer skrives etter hverandre og rollen vil vise når vi kommer til en ny kildehendelse. I første omgang kan det være naturlig å bare legge inn dåpsbarn og foreldre. Men det bør åpnes for at man senere manuelt kan legge inn faddere, dvs. nye linjer midt i tabellen eller nye kolonner. Det kan være en viktig informasjon i forbindelse med for manuell kobling.

Hvis kilden bare er scannet, men ikke transkribert, vil vi ha en tilsvarende tabell og for hver post peker til scannet bilde av side, linje eller post. Men alle/de fleste feltene vil være tomme eller fylt med en unøyaktig automatisk lesing og brukere oppfordres til å fylle inn felt når de ser på scannet bilde. Vi må ha kolonner for alle kolonnene i kilden.

Hvis kilden er en bygdebok eller for eksempel amerikansk immigrasjonsliste, vil det være mulig for en bruker å legge inn nye linjer midt i tabellen når man ønsker å lenke nye personforekomster inn i HBR-wiki. Det kan da være behov for noen andre kolonner for eksempel side- og linjenummer. Hvis det ikke er lett for en bruker å få tak i kilden, kan det være ønskelig med flere kolonner.



Figuren viser en- og toveis pekere i HBR-wiki representert med enkelt og dobbelpil. De viktigste sidene er person-, familie/husholdnings- og stedsidene. Det er pekere mellom disse sidene og til/fra kildene. Det er pekere mellom personsidene ved mulig dubletter der sidene ikke er slått sammen. Det er pekere mellom familiesidene ved at enveis peker til besteforeldre (dvs. familiesidene til foreldrene). Vi har et hierarki med stedssider. Prosjektsider internt i HBR-wiki kan ha toveis pekere til andre sider i HBR-wiki. Eksterne kilder og prosjekter kan ha toveis pekere til HBR-wiki. Hvis det blir sammenslåing av sidene i HBR-wiki, skal man automatisk flyttes pekere videre. Koblingen til lukket HBR skjer ved kopiering der 1910 folketellingen og evt. personID er referansen i begge systemene.