

Waarom Jan en Cor met elkaar trouwden

WAAROM JAN EN COR MET ELKAAR TROUWDEN

Over grote historische databestanden,
koudwatervrees en interdisciplinaire samenwerking

Rede in verkorte vorm uitgesproken bij de
aanvaarding van het ambt van bijzonder hoogleraar
Grote historische databestanden aan de Faculteit der Historische
en Kunstwetenschappen van de Erasmus Universiteit Rotterdam op
donderdag 11 juni 2009.

door Kees Mandemakers

Deze bijzondere leerstoel is gevestigd vanwege het Internationaal
Instituut voor Sociale Geschiedenis (IISG) te Amsterdam

aksant

Amsterdam

2009

ISBN 978-90-5260-352-0

© 2009, Kees Mandemakers, Almere/Amsterdam

Niets van deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie of op welke andere wijze dan ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

Omslagillustratie: Boerenbruiloft in Kruiningen. Gravure naar tekening van A. Dillens. Uit: *Le tour du monde*, 1874 (RAZ, ZG, Zel. III, III-974h).

Uitgeverij Aksant, Postbus 2169, 1000 CD Amsterdam, www.aksant.nl

*Mijnheer de Rector Magnificus,
Mijnheer de Decaan,
Geachte leden van de directie van het Internationaal Instituut voor Sociale
Geschiedenis,
Geachte toehoorders,*

Vragen, bronnen en databestanden

Welke personen wonen hun leven lang in hun geboorteplaats? Welke mensen trekken weg naar een grotere stad, een andere streek of zetten een hele grote stap en migreren naar een ander land? En komen die personen ook weer terug of blijven ze voorgoed weg? Onder welke omstandigheden groeien kinderen op en hoe groot is de kans dat het huwelijk van hun ouders wordt afgebroken en wat voor invloed heeft dat op hun kansen op maatschappelijk succes? In hoeverre valt maatschappelijk succes samen met het behalen van diploma's?

Deze en andere vragen heeft iedereen zich wel eens gesteld en zijn ook het onderwerp van wetenschappelijk onderzoek: onderzoek met een belangrijke historische dimensie. Immers, samenlevingen veranderen: waar honderd jaar geleden huwelijken vooral tot een einde kwamen door het overlijden van één van beide partners, is dit vandaag de dag vooral een gevolg van echtscheiding.¹ Onder welke omstandigheden liep men toen een verhoogd risico op vroegtijdige sterfte van een partner en werd er relatief veel of weinig hertrouwd? En met wie hertrouwde men dan? Was dit met een andere weduwe of weduwnaar of was dit met de inwonende knecht of onmisbare dienstbode? Ook de arbeidsmarkt is sterk veranderd: werkte in de negentiende eeuw bijna de helft van de Nederlanders in de landbouwsector, vandaag de dag is dit nog geen twee procent. Ook migratie zorgt voor grote veranderingen in de samenleving. Wellicht nergens komt dit binnen Nederland zo duidelijk tot uiting als in Rotterdam, een stad die inmiddels meer dan honderd verschillende nationaliteiten kent. Ook in het verleden heeft de stedelijke omgeving aan sterke veranderingen blootgestaan. Het is nog niet zo lang geleden dat een grote instroom van onder andere Duitsers, Brabanders en Zeeuwen de stad Rotterdam tot bloei heeft gebracht.²

¹ F. van Poppel, *Trouwen in Nederland. Een historisch-demografische studie van de 19^e en vroeg-20^e eeuw* (Wageningen 1992), 278. *A.A.G. Bijdragen* 33.

² Zie P. van de Laar, L. Lucassen en K. Mandemakers (red.), *Naar Rotterdam. Immigra-*

De bestudering van dit soort vragen over een lange termijn kan alleen met behulp van longitudinale gegevens. Dit zijn gegevens die worden verkregen als personen over een lange periode worden gevolgd. Onderzoek waarbij met dit type gegevens wordt gewerkt, treffen we zowel binnen de sociale wetenschappen, epidemiologie als de geschiedeniswetenschap aan onder de noemer 'levensloopbenadering'.³ Het systematisch en grootschalig volgen van levenslopen legt langetermijnprocessen bloot en maakt het mogelijk de sociale, economische en culturele factoren die veranderingen teweeg brengen te analyseren in onderlinge samenhang. De gegevens voor dit soort analyses worden opgeslagen in grootschalige longitudinale microbestanden. In Rotterdam bijvoorbeeld startte in 2003 het onderzoek *Generation R* waarin 10.000 kinderen zijn opgenomen. Deze worden vanaf de vroege zwangerschap tot hun twintigste jaar gevolgd en zullen regelmatig worden benaderd voor informatie over gezondheid, de ontwikkeling van de leefomgeving, onderwijsresultaten, etc. Centraal staat de vraag waarom het ene kind zich optimaal ontwikkelt en het andere niet.⁴

Bij historisch onderzoek is het niet meer mogelijk om personen te interviewen. Althans, als we gegevens willen verzamelen over een periode van zestig jaar of verder terug. We moeten de archieven in om informatie over de inmiddels overleden personen te achterhalen en persoonlijke en maatschappelijke factoren te vinden die de levensloop mogelijk hebben beïnvloed.

In het volgende stel ik allereerst Jan en Cor aan u voor en dan kom ik min of meer vanzelf bij de Historische Steekproef Nederlandse bevolking (HSN), en meer algemeen bij het bestaan en functie van grote historische databases. Vervolgens ga ik in op de vraag wat we ons moeten voorstellen bij 'grote historische databases'. Wat is groot, wanneer zijn ze historisch of worden ze historisch, waarom zijn ze eigenlijk belangrijk, wanneer en hoe moet je ermee werken? Dit laatste is niet onbelangrijk. De gemiddelde historicus is niet kwantitatief aangelegd, of althans hij of zij denkt dat er geen kwantitatieve aanleg is. Het werken met dit soort databestanden is in-

tie en levensloop in Rotterdam vanaf het einde van de negentiende eeuw (Aksant: Amsterdam 2006).

3 M. Kalmijn, 'Sociologische analyses van levensloopeffecten: een overzicht van economische, sociale en culturele gevolgen', *Bevolking en Gezin* 31 (2002), 3, 3-46; J. Kok, 'Transities en trajecten. De levensloopbenadering in de sociale geschiedenis', *Tijdschrift voor Sociale Geschiedenis* 26 (2000), 309-329; A.C. Liefbroer en P.A. Dykstra, *Levenslopen in verandering. Een studie naar ontwikkelingen in de levenslopen van Nederlanders geboren tussen 1900 en 1970* (SDU Uitgevers: Den Haag 2000).

4 <http://www.generationr.nl/>

derdaad niet altijd eenvoudig en – wellicht ter geruststelling – dat vinden zeker niet alleen de historici. Maar, problemen zijn er om op te lossen en daar zal het laatste deel van mijn rede overgaan.

Jan en Cor

Op afbeelding 1 ziet u de huwelijksakte van Jan Hendrik Bruning en Cornelia Stuip. Dit zijn onze Jan en Cor. Ze trouwden op 26 april 1893 te Rotterdam. Om meteen een mogelijk misverstand weg te nemen, we kunnen natuurlijk niet verklaren waarom juist zij met elkaar trouwden. Wat we wel kunnen, is patronen ontdekken in de partnerkeuze en onderzoeken of deze patronen zich in de loop van de tijd wijzigden, als gevolg van, of in samenhang met, sociale, economische, culturele, politieke of demografische veranderingen en bijgevolg in hoeverre deze patronen verschillen naar regio, land of sociale en culturele achtergrond.

Een huwelijksakte is een rijke bron van gegevens. Behalve de twee huwers staan op een huwelijksakte ook de namen vermeld van de ouders en van vier getuigen. Andere gegevens zijn de leeftijden, de beroepen en de woonplaatsen en van bruid en bruidegom ook nog de geboorteplaatsen.⁵ Van alle actoren wordt gevraagd of ze hun handtekening kunnen zetten onder de akte en dit geval hebben ze dat ook allemaal gedaan. Dit al dan niet zetten van een handtekening is voor onderzoekers van belang, omdat het een gebruikelijker indicator is bij het onderzoek naar alfabetisering.⁶

Hoe worden deze aktes nu ingezet bij onderzoek naar maatschappelijke verandering? Door middel van het beroep kunnen we zien of de twee huwers tot dezelfde sociale achtergrond behoorden of dat er sprake was van huwelijksmobiliteit. We vergelijken hiervoor dan meestal de beroepen van de vader van de bruid met die van de vader van de bruidegom. In dit geval hebben we echter meteen een probleem, want de vader van Jan is bij het

5 In dit geval betreffen de getuigen allen stadssergeanten wat duidt op een huwelijksplechtigheid zonder veel franje. Meestal zijn de getuigen familie of vrienden van het echtpaar en kunnen deze gegevens gebruikt worden voor onderzoek naar netwerken. zie bijvoorbeeld M.H.D. van Leeuwen, I. Maas en K. Mandemakers, 'Het kiezen van een huwelijkspartner in Nederland 1840-1940. De rol van de familie', in: J. Kok en M.H.D. van Leeuwen (red.), *Genegenheid en gelegenheid. Twee eeuwen partnerkeuze en huwelijk* (Aksant: Amsterdam 2005), 63-84.

6 O.W.A. Boonstra, *De waardij van eene vroege opleiding. Een onderzoek naar de implicaties van het alfabetisme op het leven van inwoners van Eindhoven en omliggende gemeenten, 1800-1920* (Wageningen 1993), 20-32. *A.A.G. Bijdragen* 34.

N^o. 424 Op heden den zesentwintigsten April achtien honderd drie en negentig ²⁴ zijn voor ons, Ambtenaar van den Burgerlijken Stand der Gemeente Rotterdam, verschenen tot het aangaan van een Huwelijk, waarvan de, bij de wet gevorderde afkondigen alhier hebben plaats gehad op den zesentwintigsten April deser maand.

Jan Hendrik Bruining oud veerentwintig jaren, lettergebruik geboren enwonende alhier, meederjarige zoon van Jan Christiaan Willem Bruining, wever, leden en Catharina Elizabeth Illoerman, zonder beroep wonende alhier.

en
Cornelia Stuij oud twintig jaren dienstbode geboren enwonende te Nieuw-Beyerland minderjarige dochter van Dirk Stuij tuisman wonende te Nieuw-Beyerland. en Mariade Graaf, overleden.

En is voor hen hierby overgelegd de huwelijke akten de de akten van den Vader des bruides en van de moeder der bruid en ten bestijf sangebanc afkondigingen te Nieuw-Beyerland.

De moeder van den Bruidegom Ende Vader der Bruid hierby tegenwoordig, hebben in dit huwelijke toegestaan en is door eerstgenoemde bevestigd verklaard dat zinde geboorte van Sharon zoon, hare voornamen abusievelijk zijn vermeld als Catharina Elizabeth, in plaats van Elisabeth Cornelia.

Daarna hebben wij hun in het openbaar afgevraagd of zij elkander aannemen tot echtgenooten en getrouwelijk al de plichten zullen vervullen, welke door de wet aan den huwelijken staat verbonden zijn; hetwelk door hen toestemmend beantwoord is. En hebben wij daarop verklaard dat zij door den echt aan elkander zijn verbonden, in tegenwoordigheid van

Abraham
Swart oud twintig jaren, Maaschel
Leegenhoek oud zestig jaren, Johannes
Christiaan Ussmann oud zieren veertig jaren
en Gysbert Wemmenen oud veeren veertig jaren
Stads Sergeanten, wonende allen alhier.

En is hiervan door ons opgemaakt deze akte, welke na voorlezing is ondertekend door ons,
de comparanten en beggetigen

J. H. Bruining C. Stuij J. C. M. J. de
E. C. Moerman D. Stuij
A. Swart M. Leegenhoek
G. Wemmenen J. Ussmann

Afbeelding 1: Huwelijksakte Rotterdam 1893, aktenummer 424, Gemeentearchief Rotterdam.

huwelijk al overleden. Maar uit een andere bron, het bevolkingsregister, weten we dat zijn vader bij leven het beroep van 'sleper' uitoefende. De schoonvader was volgens de akte 'tuinman'. Beiden vaders oefenden een ongeschoold beroep uit. In deze zin trouwde Jan binnen zijn eigen sociale groep. Jan zelf echter was letterzetter, een beroep waar een behoorlijke praktische scholing voor nodig is en een lichte sociale stijging inhield ten opzichte van zijn eigen sociale achtergrond.

Eén zwaluw maakt nog geen zomer en één huwelijksakte is nog geen bestand waar onderzoek mee gedaan kan worden. Maar, nemen we duizenden van deze aktes, dan wordt het mogelijk om onderzoek te doen naar de vraag in welke mate het gezegde 'Wie voor een dubbeltje geboren is, wordt nooit een kwartje' inderdaad opging voor het negentiende eeuwse Nederland en in hoeverre dit ook omgekeerd het geval was. En door middel van de gegevens over de geboorte- en woonplaatsen kan worden onderzocht in hoeverre Nederland zich in ruimtelijk opzicht gedurende de negentiende eeuw tot een meer geïntegreerde maatschappij heeft ontwikkeld.⁷

Met duizenden van deze aktes en andere historische bronnen komen we op het terrein van de grote historische databases. De Historische Steekproef Nederlandse bevolking (HSN) is daar een goed voorbeeld van. De HSN is gevestigd op het Internationaal Instituut voor Sociale Geschiedenis (IISG) en houdt zich sinds 1990 bezig met het aanleggen van een gegevensbestand met levensgeschiedenissen uit de negentiende en twintigste eeuw. De basis voor de HSN-database wordt gevormd door een steekproef uit de geboorteakten van de burgerlijke stand over de periode 1812-1922. De steekproef heeft een omvang van 78.000 gevallen. Dit is gemiddeld iets meer dan een half procent van de in totaal 14,5 miljoen personen die in deze periode in Nederland zijn geboren.⁸ Net als bij de huwelijksakten vinden we ook in de geboorteakten diverse gegevens die van belang zijn voor onderzoek, zoals leeftijden en beroepen van de ouders. Tevens zien we of de aangever (meestal de vader) zijn handtekening kon zetten of niet.

De gegevens voor de levenslopen zelf komen voor een groot deel uit de bevolkingsregisters. Het bevolkingsregister werd ingesteld in 1850 en

7 Zie Kok en Van Leeuwen, *Genegenheid en gelegenheid*, met daarin F. van Poppel en P. Ekamper, 'De Goudse horizon verruimd. Veranderingen in de herkomst van Goudse bruiden en bruidegoms', 181-211 en J. Kok en K. Mandemakers, 'Vrije keuze uit een beperkt aanbod. De huwelijksmarkt in Utrecht en Zeeland 1840-1940', 213-229.

8 Voor uitgebreide informatie over de HSN, zie de website: www.iisg.nl/~hsn, of K. Mandemakers, 'Historical Sample of the Netherlands', in: P. Kelly Hall, R. McCaa en G. Thorvaldsen (eds.), *Handbook of international historical microdata for population research*. (Minnesota Population Center: Minneapolis 2000), 149-178.

was tot 1940 georganiseerd op basis van het huishouden of de familie.⁹ Van gebeurtenissen zoals de geboorte van een kind, het overlijden van een gezinslid of een verhuizing naar een ander adres of gemeente moest melding worden gemaakt. Het bevolkingsregister wordt nogal eens verward met de burgerlijke stand, waar we zojuist een huwelijksakte van hebben gezien. De burgerlijke stand betreft echter een register waarin geboorten, overlijdens, huwelijken of echtscheidingen worden vastgelegd en bevestigd door middel van getuigen.¹⁰ Dit register is statisch van aard, in tegenstelling tot het bevolkingsregister dat steeds wordt bijgewerkt. Bijvoorbeeld op basis van de meldingen van geboorte en overlijden die via de burgerlijke stand binnenkomen. Ook werden bij elke nieuwe volkstelling de gegevens gecontroleerd. Dit had onder andere tot gevolg, dat flink wat personen alsnog werden doorgestreept of ingeschreven. Officieel stond op het niet melden van verhuizingen boetes, maar het is niet bekend hoe goed men daar in de negentiende eeuw de hand aan hield.

In eerste instantie werkte men met registers waarin een huishouden bij elke adresverandering opnieuw werd ingeschreven. In de loop van de tijd schakelde men, vooral in de grotere gemeenten, over op gezinskaarten, in Rotterdam al omstreeks 1890. Afbeelding 2 geeft een kopie van de gezinskaart van dezelfde Jan en Cor als uit de huwelijksakte. De gezinskaart werd aangelegd bij hun huwelijk en ze zijn redelijk honkvast, want Jan blijft in elk geval tot zijn hertrouw op 26 maart 1924 in Rotterdam wonen (hertrouw vanwege het overlijden van Cornelia in 1922).¹¹

We zien op deze gezinskaart niet alleen Jan en Cor en hun kinderen, maar ook verwanten zoals een schoonzus en zwager van Jan. Bij oudere registers en bij plaatsen buiten Rotterdam vinden we op deze kaarten ook nog inwonenden die niet verwant zijn, zoals kostgangers en knechten. Bij de HSN worden bij het volgen van de onderzoekspersoon niet alleen zijn of haar gegevens overgenomen, maar ook die van alle andere personen die zich in dezelfde gezinssituatie bevinden. Van elke persoon zien we op deze kaarten naast de naam ook nog de datum en plaats van geboorte en overlij-

9 Voor het bevolkingsregister, zie A. Knotter en A. C. Meijer (red.), *De gemeentelijke bevolkingsregisters 1850-1920* (ING: Den Haag 1995); M. Schrover en P. Vuijst, 'Familiereconstructies en bevolkingsregister', in: *Hoe krijg ik mijn bronnen aan de praat?* (VULGO: Utrecht 1990), 34-53.

10 R.F. Vulmsma, *Burgerlijke stand en bevolkingsregister*, 2e herz. druk (CBG: Den Haag 2002). CBG-reeks 10.

11 Bij de hertrouw werd er in dit geval een nieuwe gezinskaart aangelegd, wat blijkt uit de notatie 'GK II' in het vak 'overlijden' bij Jan Hendrik, dit is dus eigenlijk het verkeerde vak.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z											
VOOR- NOMMEL.	N A A M.		42822 VOORNAMEN.	GESLACHT.	RETR. TOT HET GEZIN.	GEBOORTE- TIJD.	GEBOORTE- PLAATS.	BURGELIJKE STAAT.	Kerk- genoot- schap.	BEROEP.	
1	Bruning Jan Hendrik		m	hoofd		22/10/93	Rotm	NL		H. Kerkhoven	
2	Stuij Cornelis		m	vrouw		28/9/92	Bayland	NL		"	
3	Bruning Jan Christiaan		m	27		12/12/93	Rotm	NL		"	
4	" Janus David		"	"		13/1/95	"	"		"	
5	Stuij Maria		v	bezug		14/1/94	Bayland	NL		" onderwijzer	
6	" Cornelis		m	bezug		23/11/94	"	"		" beoedebakker	
7	Bruning Hendrik Arie Adol.		m	bezug		24/1/95	Wierden	NL		" kleermaker	
8	" Johannes Hendrik		m	bezug		20/7/90	Rotm	NL		" bakker	
9	Bruning Jacob		m	bezug		31.3.91	Wierden	NL		" timmerman	
10	" Jacobus		m	bezug		27/1/95	Rotm	NL		" z.b.	
11	Bruning Jan Christiaan Hendrik		m	bezug		13-12-93	Wierden	NL		"	
12	" Jan Christiaan Hendrik		m	bezug		13/12/93	"	"		" winkelier	
	" Willem		m	bezug		13/12/93	"	"		" winkelier	

HUIZING.		Dagt- en jaar van inschrij- ving.	VOORIGE WOONPLAATS.	Dagt- en jaar van vertr.	WAARHEEN VERTROKKEN.	Dagt- en jaar van overlijden.	ONSCH- WETTO DOMICILIE	Aantekingen.
x	Oudeend 77	22/10/93	Rotm	26/1/94	Rotm	27/10/94		A.R.
x	midelste 6 Jensen	27/8/94	Rotm					
x	11. Jorden	16/1/95	Rotm	11/9/90	Rotm			
x	Mullermandhoek 8							
x	Groote eelster 72	1/10/95	Broek op Langenleijde					
x	50 1/2 25 Steune	25/1/98	Wierden	24/9/98	Wierden	17/10/98		
x	126 1/2 1/2 1/2 1/2	9/6/90	Wierden					
x		20-10-1900	Wierden	17/11/01				
x		27/1/92	Wierden	18/10/02	Wierden			
x		9-5-03	Wierden	7/2/03				
x		7-6-10	Wierden	10/1/10				
x			Wierden	21-22	Wierden			
13								wed. Tempelman Schiedamschen dijk 35

Afbeelding 2: Gezinskaart bevolkingsregister Rotterdam 1893 van Jan Hendrik Bruning en Cornelia Stuij (voorkant en achterkant). Gemeentearchief Rotterdam.

den, de burgerlijke staat, de kerkelijke gezindte, het beroep en de plaatsen waar men vandaan kwam en weer naar toe ging. Het bevolkingsregister is dus een heel rijke bron. Door koppeling van de informatie uit alle bevolkingsregisters van de meer dan duizend toen bestaande gemeenten wordt het mogelijk het volledige migratiepatroon van al onze onderzoekspersonen uit te zoeken. Op basis van de migratiegegevens en de gegevens betreffende de geboortedatum en eventuele sterfdatum kan voor elk moment in de tijd de samenstelling van het gezin en de veranderingen daarin worden bepaald.¹² In het volgende ga ik in op de wijze waarop deze gegevens voor wetenschappelijk onderzoek worden verzameld en ter beschikking gesteld.

Grote historische databases

Wat is een grote database? En wanneer is een database historisch? In de literatuur zijn niet echt goede definities voorhanden. Daar komt bij dat met de jaarlijks groeiende omvang van databases en computersystemen dit soort definities ook continu moeten worden opgerekt. Maar, er zijn wel minimale beschrijvingen van zogenaamde Very Large Databases (VLDB). We hebben het dan over databases met een omvang van minstens 1 terabyte (1000 gigabyte, dat is 1.000.000 megabyte) of met miljoenen records.¹³ Een gewone grote database zal hier qua omvang dan per definitie onderzitten, maar er zijn historische databases die ruimschoots tot de zeer grote gevallen gerekend kunnen worden. Voor het belang van een database volstaat het echter niet om alleen op de omvang te letten.

De volgende criteria zijn in de praktijk zeker zo belangrijk:¹⁴

- 1 Er wordt door meerdere personen tegelijk aan gewerkt
- 2 De database heeft een institutionele inbedding
- 3 Het gaat om forse investeringen
- 4 Er wordt gewerkt met een lange termijn strategie

12 Voor de wijze waarop de levenslopen worden gereconstrueerd, zie K. Mandemakers, 'Building life course datasets from population registers by the Historical Sample of the Netherlands (HSN)', *History and Computing* 14 (2006), 87-108.

13 http://en.wikipedia.org/wiki/Very_large_database; zie ook <http://www.vldb.org/>, website van de Very Large Data Base Endowment Inc., de non-profit organisatie ter bevordering van het onderzoek op dit gebied.

14 Zie o.a. K. Mandemakers and L. Dillon, 'Best practices with large databases on historical populations', *Historical Methods* 37 (2004), 1, 34-38; C. Harvey and J. Press, *Databases in historical research theory, methods and applications* (MacMillanPress: Houdmills etc 1996).

En de volgende criteria gelden in elk geval voor wetenschappelijke databases:

- 5 Er wordt door meerdere onderzoekers gebruik van gemaakt
- 6 De database wordt niet voor één bepaald doel of onderzoek aangelegd
- 7 De bouwers van de database zijn niet noodzakelijkerwijs dezelfde personen als de gebruikers, al is er meestal wel een band te ontdekken
- 8 De database maakt in de regel gebruik van meerdere informatiebronnen

En wanneer is een database historisch? We denken bij een historische database in eerste instantie aan een database met historisch archiefmateriaal zoals kronieken, belastingregisters, stadsrekeningen, boedelbeschrijvingen etc. Maar, we moeten ons wel realiseren dat overheid en bedrijven inmiddels al meer dan vijftig jaar gegevens elektronisch opslaan. Op termijn worden deze systemen gearchiveerd, of ze zijn al gearchiveerd, en dan hebben we een volledig elektronisch opgebouwde 'historische' database. Ook zal het onderscheid tussen historische en niet-historische databases met de tijd vager worden. Een goed opgezet administratief systeem hanteert namelijk een duidelijk onderscheid tussen historische en actuele gegevens. Daarnaast komen er steeds meer databases die een combinatie vormen van historisch bronnenmateriaal en gegevens uit hedendaagse registratiesystemen. Dit laatste kan door middel van een directe overname van data of bij privacygevoelige zaken door middel van een koppeling in een omgeving die de privacy van de gegevens waarborgt.

Om een voorbeeld te geven: Na elk overlijden worden door de Gemeentelijke Basis Administratie (GBA) persoonsgegevens, voorzover openbaar na een overlijden, ter beschikking gesteld aan het Centraal Bureau voor Genealogie (CBG). Deze gegevens worden door de HSN weer op elektronische wijze gekoppeld aan de onderzoekspersonen in de database. Een ander voorbeeld van koppeling is die met de doodsoorzakenadministratie van het Centraal Bureau voor de Statistiek (CBS). Binnen de beveiligde omgeving van het CBS worden de HSN-onderzoekspersonen aan deze administratie gekoppeld en worden er *door* het CBS voor de onderzoeker analyses uitgevoerd.

Kortom, historische databestanden lopen door tot in deze tijd en hedendaagse databestanden bevatten veel historische gegevens. Andere voorbeelden van koppelingen zijn die met geografische systemen. Door middel van coördinaten kunnen historische gegevens op oude of moderne kaarten

worden geplot.¹⁵ Het sinds kort voor het wetenschappelijk onderzoek ter beschikking gekomen register van het kadaster biedt wat dat betreft ongekende mogelijkheden.¹⁶

Heeft het dan nog wel zin, om te praten over historische databestanden? Ik meen van wel. Het gaat dan niet alleen om het historische karakter van het materiaal. Maar het gaat ook om een specifieke problematiek. Archieven zijn vaak onvolledig, niet alles is ooit opgeschreven, niet alles is bewaard en wat bewaard is, is voor een deel ook weer verloren gegaan. Het materiaal zelf is vaak moeilijk interpreteerbaar als gevolg van dubbelzinnigheden en inconsistenties, denk bijvoorbeeld aan patroniemen. In de tijd zelf was het vaak al onduidelijk wie met Cornelis Melissoon werd bedoeld, vierhonderd jaar later is dat nog een veel groter probleem.¹⁷

Historische gegevens moeten worden geïnterpreteerd om betekenis te krijgen in het wetenschappelijk onderzoek. Een voorbeeld van een dergelijke interpretatie is het indelen van personen in sociale lagen op basis van hun beroep. Op het moment dat dit soort informatie wordt toegevoegd, is er geen sprake meer van een historische database maar van een historisch informatiesysteem. Heel veel grote historische databases blijken in de praktijk ook informatiesystemen te zijn waarin meerdere lagen van gegevens aan elkaar gekoppeld worden.¹⁸ Een ander voorbeeld van verrijking is het koppelen van de geografische informatie met cartografische gegevens, we spreken dan over een Geografisch Informatie Systeem (GIS). Maar het kan veel verder gaan, ook het hele traject van data verzameling, analyse, publicatie en commentaar kan onderdeel uit gaan maken van een dergelijk systeem.¹⁹

15 I.N. Gregory and P.S. Ell, *Historical GIS. technologies, methodologies and scholarship* (Cambridge University Press: Cambridge 2007).

16 http://www.dans.knaw.nl/nl/gegevens_kadaster_snel_en_gratis_voor_universitair_onderzoek_en_onderwijs/

17 O. Boonstra, L. Breure en P. Doorn, *Past, present and future of historical information science* (NIWI-KNAW: Amsterdam 2004), 20.

18 Zie bijvoorbeeld M. Wohlgensinger, 'An historical information system methodology for analyzing power structures', *History and Computing* 13 (2001), 3, 265-282.

19 Boonstra, Breure en Doorn, *Past, present and future*, 21-23.

Micro-niveau, levensloopbenadering en aanwezigheid databases

Micro-niveau en levensloopbenadering

Gegevens kunnen op allerlei manieren worden ingedeeld, maar een belangrijk onderscheid wordt gemaakt door de eenheid waarop deze gegevens betrekking hebben. Zaken als leeftijd en geslacht hebben betrekking op een individu; maar een gemiddelde leeftijd of het percentage mannen binnen de totale bevolking heeft betrekking op een hogere geaggregeerde eenheid, dit kan een buurt zijn, of een gemeente of bedrijf, maar dit kan ook een land zijn. Gegevens op het niveau van een land worden meestal aangeduid met macro, die op het niveau van het individu met micro, alles tussen deze twee niveaus met meso. Voor databases is dit onderscheid belangrijk, omdat het richting geeft aan de wijze waarop de gegevens worden opgeslagen, gedocumenteerd, met elkaar in verband worden gebracht en uiteindelijk worden geanalyseerd. Juist vanwege het niet geaggregeerde karakter zijn het vooral de databases met microdata die een grote omvang bereiken.

In samenhang met de ontwikkeling van grote historische databases voltrok zich binnen het historische onderzoek een verandering, ook wel aangeduid met 'micro turn', waarbij in de analyse 'life histories' of levenslopen centraal kwamen te staan.²⁰ Neem de historische demografie, dat is de wetenschap die de historische ontwikkeling van de bevolking onderzoekt. Bestudeerde men eertijds de ontwikkeling van de vruchtbaarheid of mortaliteit in de vorm van gemiddelde cijfers op het niveau van landen of regio's, tegenwoordig gebeurt dit op het niveau van het individu. Dit niveau is bij uitstek geschikt om causaliteit te bestuderen.²¹ Binnen de demografie

20 J. Kok, 'Principles and prospects of the life course paradigm', *Annales de Démographie Historique* (2007), 1, 203-230.

21 Bij het onderzoek naar de verschillende tempo's in de daling van de huwelijksvruchtbaarheid in Nederland werd tot voor kort stevast uitgegaan van de gemiddelde vruchtbaarheid en de gemiddelde godsdienstige samenstelling per gemeente. De werkelijkheid is echter complexer. Binnen één gemeente kunnen er zeer grote verschillen in vruchtbaarheid bestaan. Verschillen die niet alleen afhankelijk zijn van de godsdienstige achtergrond maar ook van de sociale achtergrond van de desbetreffende individuen. Dit soort verschillen blijven verborgen bij een benadering die uitgaat van gemeentelijke gemiddeldes. Verg. O.W.A. Boonstra en A.M. van der Woude, 'Demographic transition in the Netherlands. A statistical analysis of regional differences in the level and development of the birth rate and fertility', *A.A.G. Bijdragen* 24 (1984), 1-58; Th. Engelen, 'De vruchtbaarheidskaart van Nederland', in: E. Beekink et al (red.), *Nederland in verandering. Maatschappelijke ontwikkelingen in kaart gebracht, 1800-2000* (Aksant: Amsterdam 2003), 45-70. Een voorbeeld van de nieuwe aanpak geven J. van Bavel, J. Kok en Th. Engelen, 'Hoge kinderloosheid in het interbel-

kunnen nu vragen worden beantwoord betreffende de onderlinge samenhang van bepaalde gebeurtenissen in een gezin, zoals de relatie tussen het uit huis gaan van kinderen en het overlijden van één van de ouders. Men kan processen van gezinsformatie en veranderingen volgen die perioden tot honderd jaar en zelfs meerdere generaties omvatten.²²

Nieuwe strategieën van verzamelen in combinatie met nieuwe statistische technieken hebben enorm veel mogelijkheden geopend voor onderzoek, bijvoorbeeld naar effecten uit de vroegste jeugd op de latere levensloop.²³ Ook bij het onderzoek naar migratie wordt niet meer volstaan met het verzamelen van jaarlijkse vertrekcijfers. Door middel van de archieven worden verhuizende personen opgezocht en zoveel mogelijk gevolgd in hun hele migratiepad.²⁴ In feite omvat de levensloopbenadering het hele veld van de sociale geschiedenis doordat het een kader geeft om sociale verschijnselen te bestuderen op het kruispunt van individuele en maatschappelijke verandering.²⁵

Historische databases met levenslopen, zoals de HSN, hebben het grote voordeel, dat er geen honderd jaar gewacht hoeft te worden, alvorens men de informatie over alle levenslopen compleet heeft. In het archief ligt een groot deel van de gegevens als het ware te wachten om gebruikt te worden.

lum in Nederland. De rol van godsdienst, levensstandaard en economische crisis', in: I. Maas, M.H.D. van Leeuwen en K. Mandemakers (red.), *Honderdvijftig jaar levenslopen. De Historische Steekproef Nederlandse bevolking*, boekaflevering *Mens en Maatschappij* 83 (2008), 51-80.

22 Zie bijvoorbeeld F. van Poppel, Chr. Monden en K. Mandemakers, 'Marriage timing over the generations', *Human Nature* 19 (2008), 7-22.

23 R.A. Settersten (ed.), *Invitation to the life course. Toward new understandings of later life* (Baywood Publishing: Amityville 2002). Over de invloed van de jeugd, maatschappelijke ontplooiing en omgeving op sterfte, zie M. Marmot, *Status syndrome. How your social status directly affects your health* (Bloomsbury Publishing: London 2005); G. van den Berg, M. Lindeboom and F. Portrait, 'Economic conditions early in life and individual mortality', *American Economic Review* 96 (2006), 290-302.

24 Zie bijvoorbeeld H. Bras, *Zeeuwse meiden. Dienen in de levensloop van vrouwen, ca. 1850-1950* (Aksant: Amsterdam 2002); J. Kok, K. Mandemakers & H. Wals, "'Toen scharrelde ze met haar hele zoodje naar een derde-achterkamer.'" Verhuizen als bestaansstrategie, Amsterdam 1890-1940', *Tijdschrift voor Sociale Geschiedenis* 29 (2003), 333-360 en Van Laar, Lucasen en Mandemakers, *Naar Rotterdam*.

25 G.J. Elder Jr, M. Kirkpatrick Johnson and R. Crosnoe, 'The emergence and development of life course theory', 3-19 in: J.T. Mortimer and M.J. Shanahan, *Handbook of the life course* (Plenum: New York 2003).

Aanwezigheid van microdatabases

Databases met microdata zijn er in alle maten en soorten. Voor historici zijn vooral van belang: a) databases die zijn opgeslagen in nationale data-archieven zoals in Nederland bij DANS (Data Archiving and Networked Services), b) databases waar binnen een bepaalde institutionele omgeving permanent aan wordt gewerkt zoals de HSN op het IISG en c) dynamische registraties van vooral de overheid. Dit laatste wekt misschien verbazing, maar sinds enige tijd bestaat er bij het CBS het zogenaamde Sociaal Statistische Bestand (SSB) waarin op basis van de Gemeentelijke Basis Administratie steeds meer registraties worden gekoppeld en waarmee in een omgeving die de privacy van de gegevens beschermt steeds meer onderzoek wordt gedaan.²⁶ Op termijn worden dit soort bestanden ook belangrijk voor historisch onderzoek.

De bij DANS gedeponeerde archieven betreffen in het algemeen data van een reeds afgesloten onderzoek. Door de verplichting van NWO om onderzoeksmateriaal centraal te deponeren, zal DANS ook voor historici steeds belangrijker worden. En dan gaat het niet alleen om materiaal van historici zelf, maar ook van andere wetenschappers. Het is bijvoorbeeld een kwestie van tijd dat de beoefenaren van de politieke geschiedenis op grote schaal de data gaan gebruiken die door politicologen door middel van verkiezingsonderzoeken zijn vastgelegd. Binnen de sociale wetenschappen is er een onderzoekspraktijk ontstaan waarbij datasets talloze malen worden hergebruikt door middel van zogenaamde secundaire analyses.²⁷

We kunnen historische microbestanden in eerste instantie onderscheiden naar de aard van de gegevens waarvan gebruik wordt gemaakt. We kunnen de microdata dan verdelen in statische en dynamische of longitudinale gegevens. Statisch betekent dat de gegevens alleen geldig zijn voor een bepaald moment of toestand, bijvoorbeeld het aantal kinderen in een bepaald gezin tijdens de volkstelling van 31 december 1879. Dynamisch of longitudinaal betekent dat de gegevens kunnen veranderen, bijvoorbeeld de verschillende beroepen die een persoon uitoefent over een periode van

26 Zie B. Bakker en L. Putman (red.), *De virtuele volkstelling en het Sociaal Statistisch Bestand* (CBS: Den Haag 2003); E. Schulte Nordholt, M. Hartgers en R. Gircour (eds.) *The dutch virtual census of 2001. Analysis and methodology* (Statistics Netherlands: Voorburg/Heerlen 2004).

27 Voor een mooi voorbeeld van secundaire analyse bij een onderzoek naar onderwijsmobiliteit waarbij veertien surveys uit de periode 1970-1987 zijn gecombineerd tot één dataset, zie P.M. de Graaf en H.B.G. Ganzeboom, 'Intergenerational educational mobility in the Netherlands for birth cohorts from 1891 through 1960', *The Netherlands' Journal of Social Sciences* 26 (1990), 35-50.

dertig jaar.²⁸ Binnen de longitudinale gegevens kunnen we dan weer een onderscheid maken naar de wijze waarop ze worden aangelegd. Bij een prospectieve benadering worden de personen gevolgd vanaf een bepaald moment in het leven. Bij een retrospectieve benadering worden de gegevens betreffende de levensloop achteraf verzameld door respondenten te vragen terug te kijken op hun leven.

In het volgende zal ik wat voorbeelden geven van dit soort databases. Omdat er ook gemengde vormen zijn kom ik uiteindelijk uit op vier hoofdtypen.

A Databases met statische gegevens

Databases met statische gegevens zijn veelal gebaseerd op de originele volkstellingen zoals die in een groot aantal landen nog steeds worden gehouden. De veruit belangrijkste database op dit gebied is IPUMS. IPUMS staat voor Integrated Public Use Microdata Series en is gevestigd aan de University of Minnesota, Minneapolis. Deze bevat gegevens uit de tienjaarlijkse Amerikaanse volkstellingen van 1850 tot 2000. De gegevens zijn over de jaren zoveel mogelijk vergelijkbaar gemaakt, bijvoorbeeld door eenzelfde codesysteem te maken voor alle vormen van relaties binnen een huishouden.²⁹ Het gaat hierbij steeds om 1- tot 5-procents steekproeven, uitgezonderd de volkstelling van 1880. Deze is door de leden van De Kerk van Jezus Christus van de Heiligen der Laatste Dagen (ook wel Mormonen genoemd) in de volle omvang getranscribeerd. Dit is ook gebeurd voor de volkstellingen van 1880 van Canada en het Verenigd Koninkrijk. Alle drie de tellingen zijn op systematische wijze toegankelijk gemaakt voor wetenschappelijk gebruik, samen met volkstellingen uit IJsland, Noorwegen en Zweden. Bij elkaar gaat het om meer dan 90 miljoen personen.³⁰

In Canada is de laatste jaren een vergelijkbare database gebouwd als in de Verenigde Staten onder de naam *Canadian Century Research Infrastructure* (CCRI). Opmerkelijk is de benadering die men in dit project heeft gekozen. Men kan zien, dat hier historici aan het werk zijn. Naast de cijfers

28 Voor een overzicht van de grotere historische databases met microdata, zie P. Kelly Hall, R. McCaa and G. Thorvaldsen (eds.), *Handbook of international historical microdata for population research* (Minnesota Population Center: Minneapolis 2000).

29 J.D. Hacker and C.Q. Fitch (eds.), *Building historical data infrastructure. New projects of the Minnesota population center. two special issues, Historical Methods* 36 (2003), 1, 2; zie ook <http://www.ipums.umn.edu>.

30 Het gaat hier om het North Atlantic Population Project (NAPP), zie <http://www.napp-data.org/napp/>.

zelf wordt ook veel aandacht besteed aan de historische context waarbinnen deze zijn verzameld. Wat waren de instructies waarmee de tellers op stap gingen, hoe moesten ze vast stellen wie wel en wie niet tot een bepaald huishouden behoorde, hoe moest men beroepsomschrijvingen accepteren en overnemen? Hoe veranderden deze instructies in de loop van de tijd en hoe werd er door de mensen gereageerd op het tienjaarlijks terugkerend ritueel? Wilde men wel meewerken en in welke vorm gebeurde dit? Wat stond er in de kranten op het moment van de telling? Hoe telde men de inwoners van de enorme gebieden in het Noorden, zoals de North-Western territories die nagenoeg leeg zijn? Dit spreekt niet alleen de meer traditionele historicus aan, maar geeft ook voor sociologen, demografen en economen mogelijkheden voor reflectie op het gebruik van dit historische materiaal.³¹

In Nederland vond de eerste volkstelling plaats in 1795, ten einde een kiezerslijst te kunnen opstellen voor de vergadering van de Nationale Assemblée, de wetgevende vergadering van de Bataafse Republiek. De laatste volkstelling vond plaats in 1971. Tegenwoordig kennen we een geautomatiseerde volkstelling gebaseerd op het Sociaal Statistisch Bestand van het CBS. Deze virtuele volkstelling werkt op basis van de Gemeentelijke Basis Registratie waar een groot aantal andere gegevens aan wordt gekoppeld. Voor 1960, 1971 en 2001 staan geanonimiseerde datasets ter beschikking voor wetenschappelijk onderzoek.³² Voor de periode hiervoor zijn er helaas alleen nog maar tabellen met geaggregeerde gegevens. Nederland was een van de eerste landen die een integraal bevolkingsregister invoerde en toen was het niet meer nodig de volkstellinggegevens te bewaren. Voor de tellingen van voor de invoering van dit register: 1811 (Registres Civiques),

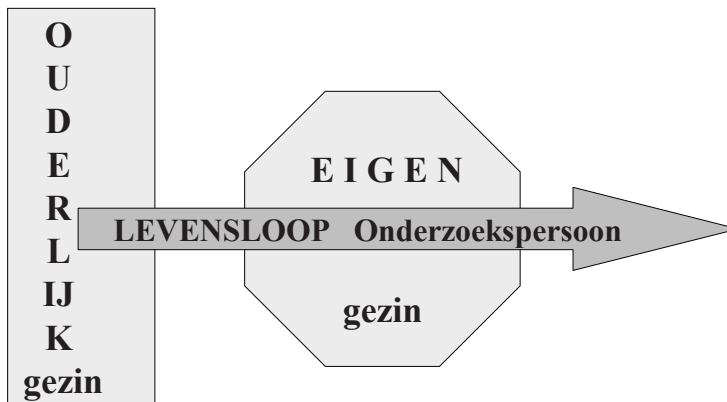
31 Het CCRI zelf concentreerde zich op de invoer en ter beschikkingstelling van de data van de vijf tellingen van de periode 1911-1951. Omdat een deel van de gegevens nog niet openbaar toegankelijk is, werd ook een speciaal netwerk ontworpen waar onderzoekers onder voorwaarden toegang toe hebben, zie C. Gaffield (ed.), *Canadian Century Research Infrastructure*, special issue *Historical Methods* 40 (2007), 2; zie ook <http://www.canada.uottawa.ca/ccri/CCRI/>; de tellingen van vòòr 1910 worden binnen andere projecten bewerkt, die van 1901 binnen *The Canadian Families Project* (special issue *Historical Methods* 33 (2000), 4.

32 P.K. Doorn en J.G.S.J. van Maarseveen, 'Inleiding. Twee eeuwen volkstelling gedigitaliseerd', in: O.W.A. Boonstra et al (red.), *Twee eeuwen Nederland geteld. Onderzoek met de digitale volks- beroeps- en woningtellingen 1795-2001* (DANS en CBS: Den Haag 2007), 3-17; J.G.S.J. van Maarseveen, 'Volkstellingen 1797-1971', in: J.G.S.J. van Maarseveen (red.), *Algemene tellingen in de twintigste eeuw* (CBS: Voorburg/Heerlen 2002), 15-88; E. Schulte Nordholt, M. Hartgers and R. Gircour (eds.) *The Dutch virtual census of 2001. Analysis and methodology* (Statistics Netherlands: Voorburg/Heerlen, 2004).

1830 en 1840 zijn op veel gemeentearchieven nog wel de originele gegevens terug te vinden.³³

B Prospectieve databases met dynamische longitudinale gegevens

Databases gebaseerd op volkstellingen geven een neerslag voor een bepaald moment. Databases gebaseerd op de bevolkingsregisters daarentegen volgen personen in principe gedurende hun gehele leven. Een voorbeeld van een dergelijke dynamische database is de al genoemde Historische Steekproef Nederlandse bevolking (HSN). Inmiddels is van bijna de helft van de in de HSN opgenomen personen de levensloop uitgezocht en ingevoerd door middel van het, door NWO gefinancierde, project *Life Courses in Context*.³⁴



Afbeelding 3: Schematisch overzicht van de levensloop van een HSN-onderzoekspersoon.

Een typische levensloop kenmerkte zich door vier verschillende situaties: a) opgroeien binnen het ouderlijk gezin, b) inwonen bij andere gezinnen of in kosthuissituaties als dienstbode, schoolleerling, ambachtsleerling, kostganger, dienstplichtige of anderszins, c) als ouder met een eigen gezin en d) inwonen als bejaarde of behoeftige, zie afbeelding 3. Echter, lang

33 De tabelgegevens van de volkstellingen zijn te downloaden, zie www.volkstellingen.nl. Voor de tellingen van 1811, zie J. L. van Zanden, *De Registres Civiques 1811 (1812, 1813)* (s-Gravenhage 1985). Broncommentaren IV.

34 Voor dit project zie K. Mandemakers, 'De Historische Steekproef Nederlandse bevolking (HSN) en het project *Life Courses in Context*', *Bevolking en Gezin* 33 (2004), 1, 91-114.

niet iedereen maakte alle fasen mee. Er was een hoge kindersterfte, niet iedereen kwam aan een huwelijk toe, of kreeg kinderen, of behaalde de leeftijd van zestig, zeventig jaar. Gedurende een leven woonde men in elk geval met veel andere personen samen. In samenhang met de 38.000 HSN onderzoekspersonen van wie de levensloop inmiddels is uitgezocht, zijn er bijna 700.000 andere personen in de database opgenomen van wie een deel van hun levensloop samenviel met die van de onderzoekspersoon.³⁵

Buiten Nederland kent men al sinds 1970 tijd vergelijkbare projecten. Wereldwijd zijn er ongeveer dertig. Zeer grote databases bestaan er in Zweden (de Demographic Database Umeå en de Scania database in Lund), de Verenigde Staten (Utah Family Database) en in Canada (Université du Québec à Chicoutimi en de PRHD te Montréal).³⁶ De database in Umeå omvat een volledige overname van de kerkelijke registers vanaf 1750 tot ongeveer 1900, voor zestig parochies verdeeld over vier regio's. Het gaat hierbij dus om een 100% invoer van alle registers over een lange periode.

Een onderzoek voor een betrekkelijk klein gebied, al dan niet door middel van een steekproef van 100%, is vrij kenmerkend voor longitudinale databases. Dit komt vooral door de praktische onmogelijkheid personen buiten een bepaalde plaats of regio te volgen. De HSN onderscheidt zich niet alleen door een klein steekproefpercentage van een half procent, maar vooral doordat de database zich over het gehele land uitstrekt. Het is tot op heden de enige database waarin voor de periode van voor 1920 het gehele migratiepad is opgenomen. Dit maakt de database van de HSN wereldwijd uniek.

C Retrospectieve databases met longitudinale gegevens

Bij de retrospectieve benadering worden personen van uit het heden teruggevolgd in het verleden. Dit gebeurt meestal door middel van interviews en/of vragenlijsten. Een voorbeeld hiervan in Nederland is de Netherlands Kinship Panel Studies (NKPS), een steekproef die in de periode 2002-2004 werd gehouden onder personen tussen de 18 en 80 jaar, dat is de geboorteperiode 1923-1984.³⁷

³⁵ Elke HSN-onderzoekspersoon treft in het bevolkingsregister tot 1940 gemiddeld dertien andere personen, via het systeem na 1940 (de persoonskaarten) en de burgerlijke stand komen er nog ongeveer vier personen bij; bij elkaar dus ongeveer achttien verschillende personen per onderzoekspersoon (deze zelf meegerekend).

³⁶ Voor een overzicht, zie de enquêtes gepubliceerd op <http://historicaldemography.net/questionnaires.php>.

³⁷ Uniek aan dit onderzoek is dat men niet alleen de onderzoekspersonen zelf interviewt maar ook familieleden. Voor meer informatie over NKPS zie <http://www.nkps.nl/>; zie ook

Een groot voordeel van een retrospectieve benadering is, dat men –binnen zekere grenzen- alle mogelijke vragen kan stellen aan de respondenten. Bij een onderzoek op basis van historische bronnen kan men wel gedragspatronen herkennen, maar er valt in de archieven niets te vinden over de redenen waarom mensen zelf denken dat ze bepaalde keuzes gemaakt hebben. Een retrospectieve aanpak heeft echter ook een aantal nadelen. In de eerste plaats is de periode die men terug kan gaan beperkt. De te interviewen personen moeten goed benaderbaar en nog in leven zijn. Daarnaast zijn er vooral bij de oudere cohorten nogal wat potentiële respondenten verdwenen als gevolg van overlijden of emigratie. Ook al omdat vandaag de dag lang niet iedereen in Nederland meer mee wil doen aan enquêtes, is het moeilijk om een min of meer representatief bestand te krijgen en, als de respondenten na een aantal jaren opnieuw worden benaderd, te behouden.

Dit alles neemt niet weg, dat alle de laatste vijftig jaar gehouden enquêtes langzaam maar zeker ook voor historici relevant worden en dit geldt des te meer voor die enquêtes die een belangrijke retrospectieve component in zich hebben.

D Gemengde typen, semi-longitudinaal

Het zal u niet verbazen dat diverse databases de verschillende benaderingen hebben willen combineren om tot een zo goed mogelijk resultaat te komen. We kennen prospectieve databases die hun retrospectieve wortels opzoeken. Een voorbeeld daarvan is de *Scottish Longitudinal Study* onder leiding van Paul Boyle.³⁸ Het gaat hier om een steekproef van 5,3% van alle inwoners van Schotland zoals te vinden in de volkstellingen van 1991. Deze personen zijn gelinkt met de volkstelling van 2001 en ze worden daarnaast gevolgd via andere registratiesystemen zoals de burgerlijke stand en medische registraties. Het retrospectieve karakter zit hem hierin dat men ook terug wil gaan zoeken, om te beginnen met de geboorteakten in de burgerlijke stand van voor 1940.

Omgekeerd zijn er ook historische databases die aansluiting zoeken bij moderne registraties. Voorbeelden daarvan zijn de Scania Database en de Utah Family Database. Zoals hiervoor betoogd, sluit ook de HSN aan op de huidige registraties met dien verstande dat het hier alleen gaat om gegevens van overleden onderzoekspersonen.

P. A. Dykstra et al, *Family Solidarity in the Netherlands* (Dutch University Press: Amsterdam 2006).

³⁸ <http://www.lscs.ac.uk/sls/>.

Een belangrijke nieuwe ontwikkeling is het dynamiseren van bestaande statistische databases. Dynamiseren houdt in, dat op zich zelf staande bronnen, zoals de volkstellingen, aan elkaar worden gekoppeld door middel van record linkage technieken. Een voorbeeld is ook weer IPUMS in de Verenigde Staten, dat de steekproeven van de volkstellingen van 1860, 1870 en 1900 koppelt aan de voor honderd procent ingevoerde volkstelling van 1880.³⁹

Het laatste voorbeeld van een semi-dynamische en meteen ook een zeer grote historische database dat ik u zal geven is LINKS. LINKS is een afkorting voor *LINKing System for historical family reconstruction* en ze omvat een reconstructie van *alle* negentiende en vroeg twintigste eeuwse families in Nederland. LINKS gaat de komende jaren gebouwd worden binnen het kader van het NWO-programma CATCH. De reconstructie is gebaseerd op de akten van de burgerlijke stand. In Nederland werken ongeveer zevenhonderd vrijwilligers aan het overnemen van de namen, woonplaatsen, leeftijden en beroepstitels uit de huwelijks-, overlijdens- en geboorteakten.⁴⁰ De huwelijksakten zijn inmiddels bijna helemaal geïndexeerd en bij de overlijdensakten is men ongeveer op de helft. Alle gegevens zijn door iedereen te raadplegen middels de websites GENLIAS en de Digitale Stamboom.⁴¹

Door middel van record linkage software zullen de huwelijksakten van kinderen, ouders en grootouders aan elkaar gekoppeld worden. En aan alle huwelijksakten worden weer de geboorteakten en overlijdensakten van de kinderen gekoppeld. Bij deze familie reconstructie zullen op basis van een

39 L. Dillon and E. Roberts, 'Introduction. Longitudinal and cross-sectional historical data: intersections and opportunities', theme issue *History and Computing* 14 (2006), 1-8; S. Ruggles, 'Linking historical censuses: A new approach', *History and Computing* 14 (2006), 213-224, zie ook http://usa.ipums.org/usa/linked_data_samples.shtml.

40 Zie <http://www.iisg.nl/hsn/news/links-project-nl.php>. Omdat de database continu wordt vernieuwd en uitgebreid, voorzien we jaarlijks een nieuwe release van LINKS met familiereconstructies, niet alleen voor wetenschappers, maar ook voor de genealogische wereld. LINKS is mogelijk dankzij een subsidie uit het NWO programma Continuous Access to Cultural Heritage (CATCH), zie http://www.nwo.nl/nwohome.nsf/pages/NWOP_5XSKYG. Het LINKS-project betreft een samenwerking tussen het IISG (Internationaal Instituut voor Sociale Geschiedenis) LIACS (Leiden Institute of Advanced Computer Science), P.J. Meertens Instituut, VKS (Virtual Knowledge Studies), NIDI (Netherlands Interdisciplinary Demographic Institute) en GENLIAS vertegenwoordigd door het Historisch Centrum Overijssel en Tresoar (Fries historisch en letterkundig centrum).

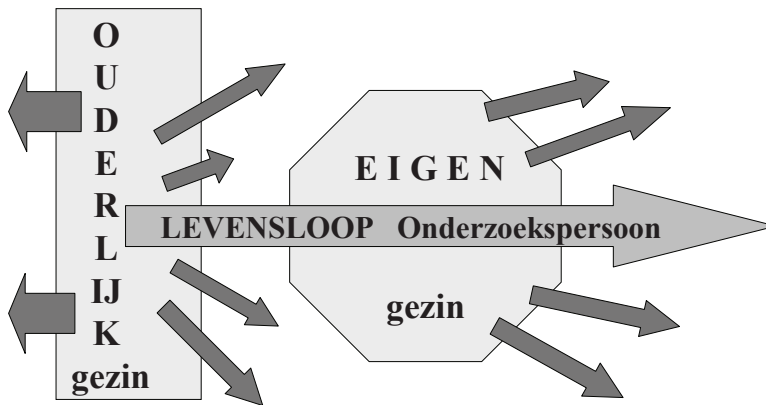
41 Zie www.genlias.nl en/of <http://www.digitalestamboom.nl/search.aspx>; er zijn verstaande plannen voor een gezamenlijke website, voorlopig nog bestaand onder de werknaam *Mijnvoorouders.nl*.

eerste schatting ongeveer twintig miljoen personen zijn betrokken die in dertig miljoen akten ongeveer honderd miljoen keer voorkomen. De beschikbaarheid van een dergelijke dataset biedt een enorm potentieel voor wetenschappelijk onderzoek. Hierbij moet niet alleen gedacht worden aan de historische demografie en de sociale en economische geschiedenis, maar ook aan naamkunde, epidemiologie, antropologie, historische sociologie en genetica.

Vanwege de hoge mate van inconsistenties in de voor- en achternamen als gevolg van fouten, verkeerde opgaven of registratie, regionale afwijkingen, fouten bij het indexeren etc is dit linken zelfs bij de burgerlijke stand niet vanzelfsprekend. Maar het LINKS-project beoogt meer, ook aanzienlijk lastiger te linken bronnen zullen in het systeem worden opgenomen, zoals adresboeken, belastingregisters en de doop- trouw- en begraafboeken van voor 1811. Ik zal hier niet teveel over uitweiden. Onderdeel van het LINKS-project is in elk geval ook de bouw van software om de binnenkomende gegevens om te zetten naar geschoonde en gestandaardiseerde eenheden en de bouw van een gevisualiseerd systeem waarmee de wetenschapper zijn benodigde data kan selecteren.

Sommigen vragen zich nu wellicht af, waarom is men ooit met de HSN begonnen als het allemaal automatisch kan? Deze conclusie gaat te snel. Binnen LINKS worden uitsluitend biologische relaties opgenomen, ongeacht de context van het huishouden en het migratiepad.⁴² Dus LINKS biedt in essentie geen migratie naar andere gemeenten, het adres waar men woonde, geen gegevens betreffende de samenstelling van het huishouden, godsdienstige achtergrond en voor een groot deel het beroep. Dit neemt niet weg dat LINKS wel een belangrijke ontwikkeling is voor de HSN-database. De in de HSN opgenomen gegevens kunnen worden gecontroleerd en aangevuld. Aanvullen gebeurt bijvoorbeeld door opname van levenloos aangegeven kinderen die alleen in overlijdensakten werden geregistreerd. Belangrijk is ook de mogelijkheid om via LINKS allerlei personen terug te vinden die ergens bij het nazoeken in de registers zijn kwijt geraakt; “tijdelijk zoek” is dan ook de benaming voor deze gevallen bij de HSN. Voor de periode van voor 1850 als er nog geen bevolkingsregister is, is LINKS van belang voor het volgen van HSN-onderzoekspersonen. Op basis van de plaats van overlijden kan er bijvoorbeeld in de registers terug worden gezocht, zodat de gegevens over huishoudens en godsdienst vanaf

42 Op basis van de ook in de index opgenomen plaatsen van geboorte- huwelijk en overlijden, kan er wel een globale indruk worden verkregen van de totale migratie zoals die uiteindelijk heeft plaats gevonden in de 19^e en vroeg 20^e eeuw.



Afbeelding 4: Schematisch overzicht van de levensloop van een HSN-onderzoekspersoon met mogelijke connecties met de LINKS-database (ouders/grootouders, broers/zussen, kinderen).

1850 ook voor eerder geboren personen in de database kunnen worden opgenomen.

Een ander belangrijk punt van LINKS is dat er meer informatie beschikbaar komt over de personen die aan de onderzoekspersoon zijn gerelateerd. Ik herinner u aan de omvang van de HSN database die tot op heden al bijna 700.000 personen telt. Afbeelding 4 geeft een schematische indruk van de levensloop van een HSN onderzoekspersoon en de mogelijke connecties met de LINKS database. We kunnen bijvoorbeeld alle huwelijksakten van de broers en zusters van de HSN onderzoekspersoon vinden.

Historici, computers en koudwatervrees?

In 1987 werd de Vereniging voor Geschiedenis en Informatica opgericht. Doelstelling van de Vereniging was het bevorderen en verspreiden van kennis met betrekking tot het gebruik van informatica bij historisch onderzoek en onderwijs, en in de erfgoedsector. En de sfeer was uitdagend, ik citeer drie toenmalige collega-bestuursleden van die vereniging: “de historicus die thans computerverwerking nog afwijst als overbodig sluit een breed gebied van onderzoeksmogelijkheden af en mag eigenlijk niet meer serieus genomen worden.” Aldus Boonstra, Breure en Doorn in hun boek

Historische Informatiekunde uit 1990, en door henzelf in 2004 nog een keer aangehaald.⁴³

En mag de historicus vandaag de dag nog serieus genomen worden? Zo'n vraag kan natuurlijk alleen in generaliserende termen beantwoord worden. In de eerste plaats is het duidelijk dat nagenoeg geen historicus nog zonder computer kan. De tekstverwerker heeft de typemachine vervangen, email voor een groot deel het brief- en telefoonverkeer, ook bij het onderwijs wordt de computer op verschillende manieren ingezet en als een historicus een grafiek maakt, gebeurt dat door middel van een spreadsheet.⁴⁴

Maar de opmerking van mijn collega's had betrekking op het onderzoek. Daarbij gaat het om de vraag of en in hoeverre historici bij hun onderzoek gebruik maken van elektronische databestanden. En verder, of historici bij het verzamelen en verwerken van gegevens wel op een technisch adequate wijze te werk gaan. En vervolgens, of historici zich bezighouden met de theoretische en praktische implicaties van de middelen die ontwikkeld worden om elektronische bestanden te ontsluiten en welke middelen er eigenlijk nodig zijn. En dan ziet het beeld er somber uit. De gemiddelde historicus staat nog even ver van de ontwikkelingen op het gebied van de historische informatiekunde als begin jaren negentig. Is dit nu een kwestie van koudwatervrees? Of ligt het subtieler? Ik denk beide. Er zijn verschillende redenen waarom een groot deel van de historici zich afzijdig houdt. In de eerste plaats zijn niet alle historici gericht op onderwerpen waarbij computerbestanden of –technieken nodig of nuttig zijn. In de tweede plaats bestuderen historici lang niet altijd *alle* aspecten van de onderwerpen waar ze mee bezig zijn. Maar als dat wel nuttig is, bijvoorbeeld omdat er voldoende gegevens zijn om bepaalde veronderstellingen toetsbaar te maken, dan gebeurt het niet of veel te weinig. Als men al weet heeft van het bestaan en de inhoud van bepaalde databestanden, dan mist men voldoende kennis om hier goed mee om te kunnen gaan.

Een oplossing voor een gebrek aan kennis is samenwerking met collega-wetenschappers die deze kennis wel hebben. Dit is eigenlijk heel gebrui-

43 L. Breure, O. Boonstra en P. Doorn (red.), *Historische Informatiekunde. Inleiding tot het gebruik van de computer bij historische studies* (Verloren Hilversum 1990), 386; geciteerd in Boonstra, Breure en Doorn, *Past, present and future*, 9.

44 Zie ook P. Doorn, *ICT en academisch onderwijs in de geestes- en maatschappijwetenschappen*, lezing Symposium *Rond ICT en Onderwijs*, ICT-centrum Faculteit der Letteren, Universiteit Utrecht, 4 februari 2004. http://www2.hum.uu.nl/solis/ict-centrum/Symposium/Symposium_februari_2004/ICT%20en%20academisch%20onderwijs.pdf.

kelijk, maar in vergelijking met veel andere wetenschappen is de doorsnee historicus tamelijk individueel bezig en sterk gericht op kwalitatieve methoden, dat wil zeggen dat er gekozen wordt voor een meer explorerende en meer beeldvormende benadering.⁴⁵ Voor een deel is dit een kwestie van belangstelling, voor een deel is dit ook een kwestie van gebrek aan kennis of aan technische mogelijkheden om voor een andere aanpak te kiezen. Ook in de gevallen dat men zelf een database opzet, gaat er een hoop mis. Ik heb de afgelopen twintig jaar heel wat historici over de vloer gehad die een database waren begonnen en er na een half jaar achter kwamen dat het niet goed ging. Of de database was niet of slecht gestructureerd, of men kreeg de data er niet meer op de gewenste manier uit. Meestal was het een gecombineerd probleem.

Veel onderzoek dat gebruik maakt van grote historische databestanden wordt uitgevoerd door sociologen, demografen en economen. Op zich is dat een goede ontwikkeling, maar het kan mijns inziens niet zo zijn, dat een historicus hele velden van onderzoek laat liggen, omdat hij niet met de methoden om kan gaan. De historicus moet zijn rol niet onderschatten. De sociale wetenschapper kan niet zonder de historicus. Het gaat daarbij om de noodzakelijke contextuele kennis van ruimte en tijd om verantwoord onderzoek te kunnen doen of om de uitkomsten van een bepaalde statistische analyse te beoordelen.

Hoe helpen we de historicus over de barrières heen? Dit kan mijns inziens op drie manieren. De eerste is een versimpeling van de wijze waarop gegevens worden aangeboden, de tweede is samenwerking tussen historici en vertegenwoordigers van andere disciplines en het derde is onderwijs. Ik ga hierna dieper in op de eerste twee strategieën. Het fundament moet hoe dan ook gelegd worden bij het onderwijs aan historici. Daar moet veel meer aandacht komen voor methoden en technieken (al is de situatie in Rotterdam overigens aanzienlijk beter dan bij de zusterfaculteiten). Juist vanwege de sterk kwalitatieve gerichtheid van het merendeel van de studenten, vergt dit een didactische aanpak waarbij van begin af aan duidelijk wordt gemaakt waarom bepaalde databestanden en technieken nuttig zijn.

Goed onderwijs in methoden en technieken is niet alleen van belang voor onderzoek met kwantitatieve gegevens. Het bestaan van grote gedigitaliseerde tekstbestanden die voor historisch onderzoek inzetbaar zijn, geeft ook het kwalitatieve onderzoek een sterk kwantitatief karakter. Het

⁴⁵ Dat is altijd al zo geweest, maar deze aanpak kreeg in de jaren tachtig wetenschapsfilosofische ondersteuning vanuit het postmodernisme ook wel aangeduid als 'linguistic turn'.

gaat dan om patroonherkenning en om toetsingsmethoden die schatten hoe groot de kans is, dat gevonden patronen niet van toevallige aard zijn.⁴⁶ De digitalisering van alle kranten bij de Koninklijke Bibliotheek zorgt ervoor dat deze automatisch doorzoekbaar worden.⁴⁷ Maar, hoe doe je dat op een goede manier? De ook in het verleden al bestaande vraag naar efficiënte zoekstrategieën, wordt nu heel actueel, alleen is het perspectief volledig omgedraaid. Kon men toen door tijdsgebrek slechts onvolledig en daarom alleen gericht zoeken, nu is het vooral de vraag hoe men op efficiënte wijze informatie krijgt uit zeer grote gegevensbestanden. Het gaat hier niet alleen om het gebruik van efficiënte zoektermen, maar ook hoe resultaten van de zoekvragen te beoordelen. Er worden thesauri en ontologieën gebouwd, om grote databestanden systematisch doorzoekbaar te maken. Om daar goed mee om te kunnen gaan, zullen onze studenten theorie en praktijk van dit soort indexen moeten leren doorgronden.

Een goed onderlegde historicus zal geavanceerde methoden en technieken niet meer negeren. Hierbij moet worden aangetekend, dat niet alle sociale wetenschappers ook alle methoden en technieken zelf beheersen. Het is niet voor niets dat juist bij de meer ingewikkelde studies combinaties van auteurs optreden: één voor de inhoudelijke kant, één voor de analyse en één voor de database zelf. Het is in de regel geen sinecure om van een bepaalde database een dataset te bouwen die geschikt is voor analyse. Vergelijk het met een proefopstelling in een laboratorium: dat kost in de regel ook meer tijd en zorg dan de uiteindelijke analyse. De samenwerking tussen onderzoekers gebeurt steeds meer in de vorm van netwerken zoals collaboratories waarop onderzoekers zijn aangesloten die wereldwijd samenwerken aan een bepaald project.⁴⁸ Op het IISG zijn momenteel verschillende van deze collaboratories actief, waaronder verschillende op basis van een gezamenlijke database.

46 Zie bijvoorbeeld J. Nerbonne, 'Crosstalk in humanities computing', *International Journal of Humanities and Arts Computing* 1 (2007), 2, 85-96; J.M. Hernández-Campoy and J.C. Conde-Silvestre, 'Computers, history and linguistics: an analysis of the social and spatial diffusion of fifteenth-century English chancery standard through computerised corpora', *History and Computing* 12 (2000), 3, 261-286.

47 http://www.kb.nl/nieuws/2006/digitalisering_kranten.html.

48 Zie T. de Moor and J.L. van Zanden, 'Do et dus (I Give So That You Give Back). Collaboratories as a new method for scholarly communication and cooperation for global history', *Historical Methods* 41 (2008), 2, 67-78.

Twee strategieën voor de beschikbaarstelling van gegevens uit historische databases

Problemen

Alvorens met oplossingen te komen, wil ik eerst nog wat dieper ingaan op de problemen die kleven aan het gebruik van microdata, gebaseerd op registraties van lang geleden. Deze waren zijn niet zo precies als we van de huidige Gemeentelijke Basis Administratie gewend zijn. En ook het GBA is lang niet altijd foutloos of compleet ondanks de zorg die er aan de gegevens wordt besteed.

De gezinskaart van Jan en Cor (afbeelding 2) was wel een heel mooi voorbeeld. In de praktijk zijn deze registers vaak aanzienlijk moeilijker te lezen. We zien doorhalingen, moeilijk leesbaar handschrift, zowel door slordig schrijven als door de vervaging die in de loop van de tijd plaats vindt. Dit geeft veel problemen bij de invoer van de gegevens. Een ander probleem is dat er voor 1900 met name in grote gemeenten eigen afwijkende systemen zijn ontwikkeld, waardoor de integratie van de registers een lastige klus is. Het kan ook zijn dat personen zich nooit hebben aangemeld in de gemeente waar ze officieel naar toe zijn gegaan. Alles bij elkaar resulteert dit voor een deel van de onderzoekspersonen in levenslopen met kleine of grote hiaten. En ontbrekende delen in de levensloop geven selectieproblemen. Dit probleem speelt voor de HSN overigens aanzienlijk minder dan bij de grote buitenlandse databases waar een vertrek naar een andere regio op zich al betekent dat men uit de database verdwijnt.

Daarnaast zijn er interpretatieproblemen met dateringen. Neem een persoon die op 1 januari 1870 als 'schoenmaker' staat ingeschreven en op 1 januari 1880 als 'leerlooier'. Wanneer veranderde hij nu van baan, hoe lang was hij al schoenmaker en hoe lang bleef hij dat?

Sommige gegevens zijn überhaupt niet gedateerd. Historische databases kennen weliswaar niet zoveel verschillende variabelen als die welke zijn gebaseerd op de hedendaagse periode, maar ze beslaan wel een zeer lange periode. En hoe langer de periode, hoe meer kans er bestaat dat bovengenoemde problemen bij het volgen van een persoon of gezin op gaan treden. Kortom vaagheid ('fuzzyness') en inconsistenties zijn schering en inslag.

Daarnaast is een longitudinale dataset, zeker vergeleken met een dataset gebouwd op basis van een volkstelling, zeer complex. Bedenk hierbij dat in de loop van een leven personen van partner veranderen, al dan niet in familieverband meerdere malen verhuizen en kinderen krijgen die zelf

ook weer komen en gaan. Bovendien werkte de toenmalige administratie zo dat bij elke verhuizing er een nieuwe inschrijving werd gemaakt. Een onderzoekspersoon kan zo wel tientallen inschrijvingen krijgen, waarbij de gegevens daarbij lang niet altijd goed overgeschreven werden. Ook is de verhuizing zelf in de regel niet gedateerd. De datum van de verhuizing moet dan geschat worden op basis van andere wel gedateerde gebeurtenissen in de inschrijvingen op de adressen waartussen verhuisd wordt.⁴⁹

Alle grote databases met longitudinale gegevens kampen met dit probleem van complexiteit en bij onoordeelkundig gebruik kan er veel fout gaan. Sommige databases weigeren daarom zelfs met algemene releases te komen waaruit onderzoekers kunnen putten. Bij de Demographic Database in Umeå bijvoorbeeld zijn er speciale programmeurs die met onderzoekers de vragen doornemen en vervolgens een voor het desbetreffende onderzoek specifieke dataset maken, waarbij knopen worden doorgelicht hoe gegevens moeten worden geïnterpreteerd, gedateerd, gebruikt of niet gebruikt. Maar dit is een kostbare situatie en weinig flexibel, daar de onderzoekers naar het instituut in Umeå moeten komen voor het analyseren van de data.

Platte databestanden

De strategie van versimpeling werkt op verschillende manieren. Er wordt alleen een deel van de longitudinale gegevens aangeboden en dit gebeurt in de vorm van een zogenaamde platte data structuur. Dit betekent dat de rijkdom van de dataset slechts zeer gedeeltelijk wordt doorgegeven, maar dat is ook lang niet altijd nodig. Een simpele dataset bevat bijvoorbeeld alleen een eerste huwelijksdatum en het aantal keren dat een persoon getrouwd geweest is en geeft geen aparte tabel voor alle gegevens van alle mogelijke huwelijken die een persoon gehad kan hebben.⁵⁰ Een vereenvoudiging houdt ook in, dat mogelijke keuzes en onzekerheden voor de onderzoeker worden opgelost en dat alle gegevens zoveel mogelijk worden gestandaardiseerd. Een platte dataset kan gemakkelijk in een spreadsheet worden geïmporteerd, dit is voor historici het meest gebruikte programma om met kwantitatieve gegevens te werken. De bij de database behorende documentatie wordt per variabele aangeleverd en is beperkt van inhoud.

49 Voor een systematische benadering, zie het Amsterdam protocol, Mandemakers and Dillon, 'Best practices'.

50 Normaliter wordt er binnen een database voor huwelijken (al dan niet in combinatie met andere dynamische gegevens) een aparte tabel gemaakt.

Wil de historicus of andere gebruiker meer weten over de data, dan biedt deze wel een directe ingang op de volledige documentatie.

Op het IISG wordt momenteel gewerkt aan de zogenaamde HSN-datamachine met daarin de belangrijkste gegevens uit de HSN-database. Deze zullen via de website van de HSN op interactieve wijze in een plat databestand worden aangeboden. De gebruiker selecteert die variabelen die hij nodig heeft en krijgt die in een bestand thuis gestuurd. We verwachten aan het einde van het jaar met een eerste release te zullen komen. Deze strategie van versimpeling wordt ondersteund door het grote succes van databases die zijn gebaseerd op volkstellingen. Ik denk hierbij aan het al eerder genoemde IPUMS van het Minnesota Population Center waar onder leiding van Steve Ruggles gigantische investeringen ook een vloed aan publicaties hebben opgeleverd van sociale wetenschappers en historici. Een belangrijke reden voor dit succes is de simpele structuur waarin de data worden aangeboden.

De tweede strategie is samenwerking tussen verschillende wetenschappers, disciplines en instituten. Dit is volgens mij de enige reële manier om de grote longitudinale databestanden volledig en op efficiënte wijze te benutten. Hier zal het laatste deel van mijn rede overgaan.

Internationale samenwerking

Het bestaan van databases in verschillende landen voor verschillende tijden nodigt uit tot vergelijkend onderzoek. Een pionier op dit gebied was het *Eurasia project*. In dit project werden de levenslopen onderzocht van historische bevolkingen van België, Zweden, Italië, Japan en China. In *Life Under Pressure* presenteerde de groep de eerste resultaten. Op basis van een vergelijkende analyse konden bijvoorbeeld verschillen in sterfte worden verklaard uit machtsverschillen tussen generaties en sexen op basis waarvan binnen families het voedsel werd verdeeld of werd gereageerd op economische crises.⁵¹

Bij het project kwamen echter ook veel problemen naar voren die samenhangen met verschillen in de betekenis van de gegevens, de wijze waarop met onbekende gegevens of ontbrekende dateringen moest worden omgegaan en last but not least de manier waarop de data waren gestructureerd en gedocumenteerd. Dit verschilde per database. Uiteindelijk moest men er vanaf zien om met één gezamenlijke datastructuur te werken. En dat

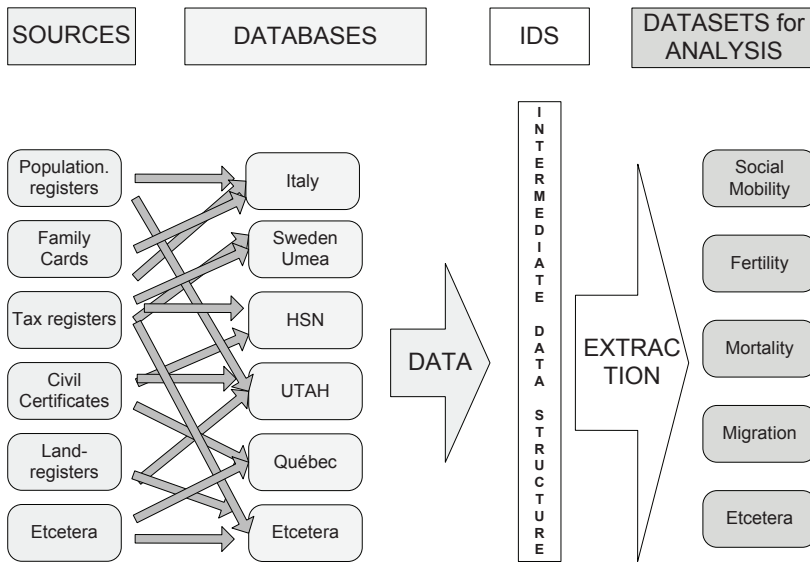
⁵¹ T. Bengtsson, C. Campbell and J.Z. Lee, *Life under pressure. Mortality and living standards in Europe and Asia, 1700-1900* (The MIT Press: Cambridge Massachusetts London 2004).

waren dan nog maar vijf databases. Zoals aan het begin van deze rede reeds vermeld, zijn er inmiddels meer dan dertig grote historische databases met longitudinale microdata. De uitdaging om deze databases op één lijn te krijgen spreekt voor zich, maar alleen al het aantal maakt de poging er op het eerste gezicht niet gemakkelijker op. Daarnaast is er het probleem dat de behoefte aan data, qua inhoud, mate van exactheid en structuur, varieert met de aard van de wetenschappelijke analyse. Het inzicht dat het onmogelijk is om één ideale dataset te maken voor alle onderzoekers was bij de HSN al langer aanwezig.

In 2006 nam de HSN de leiding in de discussie hoe tot een oplossing te komen voor de zojuist geschetste problematiek. Mede gefinancierd door NWO-Geesteswetenschappen is er veel overleg geweest en is er een consensus ontstaan over de wijze waarop deze problematiek moet worden opgelost.⁵² In essentie komt het erop neer dat de data van de verschillende databases op een tamelijk basaal niveau in een gemeenschappelijke datastructuur worden ondergebracht: de Intermediate Data Structure (IDS). Op basis van deze IDS worden er applicaties ontwikkeld die de data omzetten in een structuur die voor een bepaald onderzoek gewenst is. Zowel de IDS als de onderzoeksspecifieke software worden gedeeld en verspreid door middel van een gezamenlijk onderhouden collaboratory.

Afbeelding 5 geeft een schematisch overzicht. Links staan de databases die per regio of land verschillende bronnen omvatten. Alle gegevens worden vervolgens op een systematische manier in één en dezelfde dataset ondergebracht, de zogenaamde Intermediate Data Structure (IDS). Door middel van standaardisering worden in deze stap meteen de gemeenschappelijke variabelen op dezelfde wijze gestructureerd en gedocumenteerd. Vervolgens worden er tools gebouwd die datasets produceren afhankelijk van de aard van het te verrichten onderzoek. Zo heeft een onderzoek naar sociale mobiliteit andere gegevens nodig dan een onderzoek naar vruchtbaarheid. Ook kan men met minder nauwkeurige dateringen volstaan. In de praktijk betekent dit dat, afhankelijk van het onderzoek, verschillende delen van de dataset tot een analysebestand gemaakt worden. Het bestaan

52 'Towards a global history of life courses. Creating a network for the development of data structures for standardized longitudinal historical data', granted by ICPSR Ann Arbor, DDB Umeå and Netherlands Organisation for Scientific Research (NWO), Humanities (Internationalizing, 236-53-004). Een beschrijving van de IDS verschijnt als G. Alter, K. Mandemakers and M. Gutmann, 'Defining and distributing longitudinal historical data in a general way through an intermediate structure', *Historical Social Research*, to be published August 2009.



Afbeelding 5: Schema van de dataflow door middel van een Intermediate Data Structure (IDS) als middel voor vergelijkend wetenschappelijk onderzoek van historische longitudinale databases.

van een dergelijke tool betekent ook dat een historicus die onderzoek gaat doen naar sociale mobiliteit voor elke database met geschikte data met een druk op de knop een nieuwe dataset kan bouwen. De historicus kan zich dan volledig richten op het bestuderen van de documentatie behorend bij het databestand. En ook deze is voor alle databases op een gestandaardiseerde wijze ingericht.

Samenvattend, in plaats van een moeizame studie van de structuur en inhoud van de verschillende databases en de bouw van verschillende stukken software om de data in een analyseset te krijgen, kan er nu worden volstaan met het maken van één tool die werkt op basis van de IDS. Het blijft voor de onderzoeker bovendien altijd mogelijk om zelf zonder gebruikmaking van bepaalde software op basis van de IDS een eigen heel specifieke dataset te bouwen.

Er is intussen een werkend voorbeeld van een tool ontwikkeld door George Alter (ICPSR) voor de bouw van een dataset voor onderzoek naar vruchtbaarheid. En bij verschillende databases is men begonnen de data te converteren naar de IDS (onder andere bij DDB Umeå en bij de Scania database in Lund). De snelle overname van deze structuur door deze databases komt mede, omdat de systematiek van de IDS uitermate geschikt is

om ook de eigen bouw van software en documentatie op een systematische manier aan te pakken.

Grote historische databestanden en de FHK

Reeds vanaf de oprichting van de Faculteit der Historische en Kunstwetenschappen bij de Erasmus Universiteit, in de vorm van de toenmalige subfaculteit Maatschappijgeschiedenis, heeft de bestudering van de samenleving er centraal gestaan in het onderwijs en onderzoek. Eén van de problemen waar men aanvankelijk mee te kampen had, was het ontbreken van geschikte historische gegevens waarmee de maatschappelijke ontwikkeling op het niveau van individuen onderzocht konden worden.

De “elektronische revolutie” heeft daar verandering in gebracht. De opkomst van computers waarmee grote hoeveelheden data kunnen worden opgeslagen en bewerkt heeft geheel nieuwe vormen van sociaal-historisch en historisch-demografisch onderzoek mogelijk gemaakt. Zoals ik heb laten zien, bestaan er in binnen- en buitenland inmiddels zeer grote databestanden met gegevens op microniveau, dat wil zeggen op het niveau van individuen, en deze bestanden worden wetenschappelijk steeds belangrijker. Deze ontwikkeling roept geheel nieuwe theoretische, methodologische en onderzoekspraktische vragen op. Het Internationaal Instituut voor Sociale Geschiedenis (KNAW), waar de HSN is ondergebracht, acht het van groot belang dat aan deze vragen stelselmatig aandacht wordt geschonken.

Het is mijn doel om door middel van de bijzondere leerstoel ‘Grote historische databestanden’ aanstaande historici vertrouwd te maken met de inhoud en structuur van deze databases en ze in te wijden in de wijze waarop met deze gegevens wordt gewerkt. Ik zie deze leerstoel ook als een middel om het hierboven geschetste internationale netwerk van historische databases met microdata te verstevigen en te verankeren. Ik verheug me erop studenten en promovendi te interesseren in onderzoek met grote databestanden. Het onderzoek zal zich richten op zowel inhoudelijke als op datatechnische zaken. De inhoudelijke vragen komen vanuit het perspectief van het onderzoek naar levenslopen. Bij datatechnische zaken gaat het om onderzoek naar de optimale structuur van deze databases voor zowel opslag als terbeschikkingstelling voor onderzoek en naar de mogelijkheden om de data op systematische wijze te ‘verbeteren’, bijvoorbeeld door onderzoek naar schattingsprocedures voor ontbrekende dateringen. Samenvattend wil ik met deze leerstoel het belang en de problematiek van

de grote historische databestanden meer systematisch onder de aandacht brengen dan tot op heden is gebeurd.

Mijnheer de rector, dames en heren,

Aan het eind van deze rede gekomen, past een woord van dank aan de directie van het Internationaal Instituut voor Sociale Geschiedenis, de Faculteit der Historische en Kunstwetenschappen, het Curatorium van deze leerstoel en het College van Bestuur van de Erasmus Universiteit Rotterdam. Ik ben hen erkentelijk voor de instelling van deze leerstoel en het in mij gestelde vertrouwen.

In 1981 studeerde ik af bij Theo van Tijn en eigenlijk ook bij Jan Lucasen en Eric Fischer. Zonder de andere medewerkers van de afdeling sociale en economische geschiedenis in Utrecht tekort te willen doen, zijn zij het vooral die me tot wetenschapper hebben gevormd. In juni 1981 kwam ik in Rotterdam bij de toenmalige subfaculteit maatschappijgeschiedenis i.o. voor promotieonderzoek naar de sociale achtergrond van leerlingen bij het secundaire onderwijs. Mijn promotor Henk van Dijk was een van de eerste historici die het belang van de PC inzag en mij daarop introduceerde. Mijn andere promotor, Jaap Dronkers, leerde ik kennen toen ik lid werd van de SISWO-werkgroep Longitudinale Analyse. We schreven samen het rapport over de noodzaak van de 'Geboortedatabank', dat uiteindelijk de Historische Steekproef Nederlandse bevolking (HSN) werd. Het idee voor die steekproef had ik opgedaan bij mijn onderzoek, waarbij ik het hele land doorreisde om de archieven van ongeveer 150 scholen te bezoeken. In de trein had ik alle tijd om me af te vragen waarom sociologen hun dataset altijd gewoon kregen en historici hun data zelf moesten verzamelen. Mijn eigen dataverzameling kostte zeker veel tijd, maar Henk van Dijk verloor nooit het vertrouwen in de goede afloop.

Inmiddels zijn we bij de HSN beland, tussentijds was ik ook één jaar docent Historische Informatiekunde aan de EUR en daarna in deeltijd tien jaar aan de Vrije Universiteit. Eric Akkerman was een goede collega en achterwacht als dat nodig was. Harry Ganzeboom werd voorzitter van de stichting HSN, ik werd secretaris, en samen gingen we op pad om onderdak te vinden. Op het IISG werden we ontvangen door Jaap Kloosterman en Henk Wals. Op hun advies deden we een aanvraag bij het Ministerie van Onderwijs & Wetenschappen voor een proefproject in de provincie Utrecht. Dat werd toegekend en vanaf 1991 is de HSN onder de hoede van

het IISG, waar we zijn gegroeid tot wat we nu zijn en waar de directie ook in moeilijke tijden de HSN bleef steunen. Het leiden van een dergelijke onderneming is niet altijd even gemakkelijk en ik heb dat ook zeker niet alleen gedaan. Ik dank Jan Kok, Lex Heerma van Voss en Karin Hofmeester met wie allerlei technische en tactische kwesties konden worden doorgesproken. Ook de leden van de stuurgroep van de HSN zijn belangrijk geweest in de ontwikkeling van mijzelf en de HSN. Ik denk dan met name aan Frans van Poppel en Hans Knippenberg. Het was de laatste die me einde 1990 overhaalde om de HSN te gaan trekken.

De HSN heeft de afgelopen twintig jaar welgeteld 135 verschillende personen in dienst gehad, een vijftigtal die via uitzend- en invoerbureaus werkten niet meegerekend. Het leiden van deze organisatie werd in eerste instantie samen gedaan met Jan Kok en later met Sjaak van der Velden, Jaap Bording en Evelien Walhout. Was ik nogal eens kort door de bocht, ze wisten me toch steeds uit te leggen dat ook de HSN mensenwerk is en dat je bij zo'n ingewikkelde onderneming niet kunt verwachten dat alles meteen volgens plan verloopt. De HSN volgt een half procent van alle in de negentiende en vroegtwintigste eeuw geboren personen. Dit betekent dat alle bestaande bevolkingsregisters (meer dan 50.000 afzonderlijke delen) en alle registers van de burgerlijke stand wel een keer geraadpleegd moesten worden. Het is dan ook een enorme logistieke operatie waarbij een goede samenwerking met het archiefwezen onontbeerlijk is. Zonder de zeer goed gestructureerde archiefdienst en hulp van de medewerkers zou het bouwen van een database als de HSN onmogelijk zijn geweest.

Zoals gezegd werkte ik in de jaren tachtig bij dezelfde faculteit waar ik nu opnieuw ben aangesteld en ik voel me er ook meteen weer thuis. Niet alleen door de hernieuwde kennismaking met talrijke oud-collega's, ik deel de kamer met Geert Stevens, maar ook door de warme ontvangst door mijn 'nieuwe' collega's, met name Hein Klemann, Maria Grever en Dick Douwes. Sinds mijn vertrek aan de Vrije Universiteit heb ik geen onderwijs meer gegeven en dit en het contact met de studenten heb ik gemist. Ik hoop dit de komende jaren ruimschoots in te halen.

Tenslotte, het verheugt me ontzettend dat mijn ouders, Han en Riet, er vandaag bij kunnen zijn. Het belang van een goede opleiding werd er vanaf het begin af aan door hen in gehamerd en met de wortel en de stok werd ik door de lagere en middelbare school geleid. Ze gaven me ook alle vertrouwen, toen ik op zeventienjarige leeftijd naar de universiteit van Utrecht trok en daar ben ik ze dankbaar voor. De HSN standaardlevensloop onge-

veer volgend ben ik op bijna 26-jarige leeftijd getrouwd met Frouke Hansum, we kregen twee kinderen, Jornt en Myrthe en we zijn – iets minder standaard – nog steeds bij elkaar. Zonder haar zou mijn levensloop er heel anders uit hebben gezien. De twee meest aantrekkelijke aspecten van de wetenschapsbeoefening, namelijk het in elkaar overgaan van werk en liefhebberij en het steeds weer staan voor nieuwe uitdagingen, is risicovol voor een goede werkplanning en daarmee voor het gezinsleven. Ik ben Frouke dankbaar dat ze deze tweede liefde van mij geaccepteerd heeft en al vanaf het handmatig sorteren van kaartenbakken vol met Waalwijkse schoenmakers tot en met de wereldwijde Intermediate Data Structure actief en vol belangstelling aanwezig is.

Ik heb gezegd.