



HAL
open science

Rapport om Samfundsvidenskabelige Primærdata

Barfort Sten, Anders Sparre Conrad, Bertil Dorch, Katja Guldbæk, Asger Væring Larsen, Simone Schipp von Branitz Nielsen, Tina B Pipa, Jens Ludvigsen, Sigfrid Lundberg

► **To cite this version:**

Barfort Sten, Anders Sparre Conrad, Bertil Dorch, Katja Guldbæk, Asger Væring Larsen, et al.. Rapport om Samfundsvidenskabelige Primærdata. 2009. hprints-00451000

HAL Id: hprints-00451000

<https://hal-hprints.archives-ouvertes.fr/hprints-00451000v1>

Preprint submitted on 27 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Samfundsvidenskabelige Primærdata

af

Projektgruppe bestående af: Sten Barfort¹, Anders Sparre Conrad⁵, Søren Bertil Fabricius Dorch², Katja Guldbæk⁵, Asger Væring Larsen², Jens Ludvigsen³, Sigfrid Lundberg⁴, Simone Schipp von Branitz Nielsen⁵ & Tina B. Pipa¹

¹ Det Samfundsvidenskabelige Fakultetsbibliotek, Det Kongelige Bibliotek / Københavns Universitets Biblioteks- og Informationsservice

² KUBIS Videncenter for Videnskabelig Kommunikation, Det Kongelige Bibliotek / Københavns Universitets Biblioteks- og Informationsservice

³ Institut for Sociologi, Københavns Universitet

⁴ Digital Infrastruktur og Service, Det Kongelige Bibliotek

⁵ Institut for Psykologi, Københavns Universitet

Rapport om samfundsvidenskabelige primærdata

Indhold

1. Opsummering	3
2. Indledning	4
3. Projektets hypoteser, hovedspørgsmål og ledetråde	5
<i>Primære hypoteser</i>	5
<i>Hovedspørgsmål</i>	5
<i>Sekundære hypoteser</i>	6
<i>Ledetråde for interviews</i>	6
4. Litteraturstudium	8
<i>Bevaring</i>	8
<i>Regler</i>	8
<i>Genbrug af data</i>	9
<i>Værktøj</i>	10
<i>Forskningsmetoder</i>	10
<i>Forskerservice</i>	11
<i>Fordele ved dataregistrering og dataarkivering</i>	11
5. Praksisstudium	13
<i>Dansk Dataarkiv</i>	13
<i>Sociologisk Databank</i>	14
6. Interviews	16
<i>Interview 1</i>	16
<i>Interview 2</i>	16
<i>Interview 3</i>	17
<i>Interview 4</i>	17
<i>Interview 5</i>	17
<i>Interview 6</i>	18
<i>Interview 7</i>	18
<i>Interview 8</i>	19
<i>Interview 9</i>	19
<i>Opsummering på tværs</i>	20
<i>Relation til primære og sekundære hypoteser</i>	21
7. Konklusion og anbefaling	22
Bilag	23
A. <i>Deltagere i projektet</i>	23
B. <i>Litteraturliste</i>	24
C. <i>Interviewguide</i>	25
D. <i>Informanter</i>	30

1. Opsummering

Nærværende rapport er resultat af et KUBIS-projekt, om videnskabelige primærdata inden for samfundsvidenskaberne, der forløb i perioden marts til december 2009. Projekt er støttet med midler fra KUBIS' Innovationspulje 2009.

Projektets udgangspunkt er et ønske fra KUBIS, om at undersøge mulighederne for at bidrage med biblioteksmæssige kompetencer til arbejdet med videnskabelige digitale primærdata. Det vil sige, om bibliotekernes kompetencer og ressourcer indenfor indsamling, registrering, digitalisering, arkivering, bevaring og formidling kan sættes aktivt i spil, med det formål at udvikle services til forbedring af den del af de videnskabelige kommunikationsprocesser, der er forbundet med anvendelsen af videnskabelige primærdata.

På det samfundsvidenskabelige område er primærdata mangfoldige – både med hensyn til indhold, anvendelse, mængde og formater. I denne rapport forsøges det, gennem en beskrivelse af det forløbne projekt inklusive en opsummering og konklusion, at belyse universitetsbibliotekets potentielle rolle med hensyn til videnskabelige primærdata.

Undersøgelserne i projektet viser, at de elektroniske data, der indsamles og opbevares lokalt af forskere, især stammer fra spørgeskemaer, tests, interviews og observationer. Hovedparten af data har en kvalitativ karakter. Data opbevares på PC'er eller fællesdrev – og i mange tilfælde uden effektiv eller velorganiseret backup. En del data er personfølsomme og nogle data indsamles og/eller opbevares af eller i samarbejde med eksterne institutioner, eksempelvis Dansk Data Arkiv.

Den lokale behandling af data sker gennem anvendelse af mangeartede software – og support i forbindelse med disse programmer er enten i høj grad overladt til forskerne selv, eller præget af tilfældige personlige relationer; hele processen forekommer uigennemskuelig og uoverkommelig for flere forskere.

Nogle forskere udveksler og deler kun bearbejdede data, mens andre også udveksler og deler rådata. Udvekslingen sker normalt via e-mails – hvis der ikke er tale om direkte download fra en central institution. Virtuelle netværk benyttes ikke eller kun lidt og egentlig e-science er der (endnu) ikke tale om.



Bedre og mere velorganiseret opbevaring af data kan bidrage til øget effektivitet i forskningsprocessen, øget synlighed af aktuel og tidligere forskning og øget genbrug af empiriske data. I mindre grad er der tegn på nye forskningsformer og tværvidenskabelig forskning



Der er problemer mht. bevaring af data, forskerne er opmærksomme på regler for anvendelsen af data og på at genbrug af egne data er nyttig. Adgang til andres data sker ikke på en systematisk og velorganiseret måde, og der er et behov for information om programmer – der er behov for øget forskerservice på nogle af disse områder

Konklusion: De samfundsvidenskabelige forskere har brug for opbevaring af data et sikkert sted, med adgang over længere tid. Set i relation til bibliotekets kompetencer forekommer det oplagt at sætte konkrete pilotprojekter i værk på området "bevaring og personlig adgang til data"; med andre ord, et forsknings "myspace", hvorpå fremtidige *overlay services* kan bygges.

1. Indledning

Projektet blev igangsat af KUBIS i marts 2009, med det formål at undersøge hvilke primærdata der produceres og anvendes af forskere, indenfor forskellige samfundsvidenskaber, hvordan disse data opbevares og bearbejdes, hvilke mængder og former for data der er tale om, i hvilken udstrækning og på hvilke måder disse data deles mellem forskellige forskere, og hvilke behov forskere har i forbindelse med anvendelsen af primærdata til ny forskning, undervisning og studenteropgaver.

Projektet har desuden til formål at belyse universitetsbibliotekets muligheder for at understøtte forskere i deres arbejde med primærdata og for at bidrage til formidling af primærdata til forskere, undervisere og studerende. Antagelsen er, at bibliotekets kernekompetencer hvad angår for eksempel indsamling, registrering, digitalisering, arkivering og formidling med fordel kan anvendes på området – med henblik på udvikling af services til forbedring af den del af den videnskabelige kommunikation, som vedrører anvendelsen af primærdata.

Projektet er oprindeligt tænkt som bestående af forskellige faser eller dele. Denne del, der præsenteres i de følgende afsnit, er tænkt som et opklarende, eksplorativt projekt, der har til formål at belyse ovennævnte spørgsmål på tre forskellige måder (via tre forskellige *spor*) – dels gennem interviews af forskere ved Københavns Universitet (KU), dels gennem studier af litteratur som omhandler forholdet mellem primærdata, forskningsprocesser og IT indenfor forskellige videnskaber – det vil sige, som vedrører forskellige aspekter af e-science – og dels gennem studier af hvordan henholdsvis Dansk Data Arkiv (DDA) og Sociologisk Databank (SD) indsamler, opbevarer og formidler samfundsvidenskabelige primærdata.



Denne del af projektet skal afklare behovet for et eventuelt opfølgende pilotprojekt vedrørende samfundsvidenskabelige primærdata og det konkrete indhold i et sådant praktisk pilotprojekt.

Forskellige personer har haft det primære ansvar for at beskrive og opsummere de forskellige dele af projektet. Centerleder Bertil Dorch har som projektleder beskrevet projektets hypoteser og hovedspørgsmål m.m., informationsspecialist Asger V. Larsen har opsummeret litteraturstudierne, bibliotekar Jens Ludvigsen har skrevet om Dansk Data Arkiv og Sociologisk Databank, og informationsspecialist Sten Barfort har opsummeret interviews og sammenskrevet de enkelte dele. En oversigt over alle personer, som har været involveret i denne første del af projektet, findes som bilag (jf. Bilag A).

3. Projektets hypoteser, hovedspørgsmål og ledetråde.

Her oplystes hypoteser og ledetråde til brug for analysedelen i projektets tre spor: Interviews (empiri), litteraturstudium og praksisstudium. Foruden ledetråde for interviews findes en egentlig interviewguide – denne vedlægges som bilag (jf. Bilag C).

De primære hypoteser udgør projektets "research questions": Det vil sige, disse hypoteser repræsenterer de overordnede antagelser, der gør det relevant for forskningsbiblioteker som KUBIS, at beskæftige sig med videnskabelige primærdata, som informationsressourcer der bør indgå i bibliotekets formidling til universitetet. Hypoteserne fokuserer på relevansen af bibliotekets kompetencer i forhold til at skabe værdig for forskningen, samt på eksistensen af – og muligheden for at arbejde med primærdata.

Primære hypoteser

1. Bibliotekets kompetencer (eksempelvis professionel indsamling, registrering, systematisering, digitalisering, arkivering og formidling) er relevante for arbejdsgangen omkring videnskabelige primærdata: Disse kompetencer har værdi for forskeren, da de kan bidrage til øget effektivitet af forskningsprocessen (for eksempel tids- og kvalitetsmæssigt), øget synlighed af både aktuel og tidligere forskning, facilitering af nye forskningsformer og tværvideenskabelig forskning, samt øget genbrug af empiriske data til nye forskningsprocesser.
2. Der eksisterer elektroniske videnskabelige primærdata indenfor samfundsvidenskaberne, og der findes et behov for en ny form for struktur og nye teknologier i forhold til anvendelsen af disse data. Dette behov er ikke opfyldt af allerede eksisterende systemer og praksis – som er beregnet på og udviklet til tidligere tiders dataformater (for eksempel papirmaterialer og fotografier) og empiri.
3. Det er muligt at beskrive arbejdsgange (workflows) for arbejdet med samfundsvidenskabelige primærdata. Baseret på brugerbehovet kan bibliotekets kompetencer anvendes i disse workflows.

Hovedspørgsmålene i dette projekt er en analytisk operationalisering af de primære hypoteser; det vil sige, med udgangspunkt i hypoteserne, stilles disse spørgsmål til den indsamlede empiri, for – i sidste ende – at muliggøre en falsificering af hypoteserne.

Hovedspørgsmål

1. Hvad er bibliotekets potentielle rolle mht. opsamling og formidling af data? Det vil sige, hvor og hvordan kan bibliotekernes kompetencer sættes konkret ind i opbygning af en relevant infrastruktur indenfor samfundsvidenskaberne?
2. Hvad og hvordan er dataomfanget (hvor meget, hvad, hvordan)?
3. Hvad og hvordan er brugerbehovet?
4. Hvordan er brugeradfærden?

5. Hvad er erfaringerne fra andre steder?
6. Er der mindst to datasæt fra konkrete samfundsvidenskabelige projekter på KU, hvor data ikke er tilgængelige via projektets publikationer – eller på anden vis?

Som praktisk supplement til de primære hypoteser og hovedspørgsmålene, har vi konstrueret et sæt af emneinddelte sekundære hypoteser, der muliggør fokus på særligt biblioteks-relevante aspekter af forskernes anvendelse af videnskabelige primærdata.

Sekundære hypoteser

1. Bevaring
 - Data forsvinder eller er udsat for utilstrækkelig bevaring
2. Regler
 - Forskerne er opmærksomme på datatilsynets og lovgivningens regler samt etiske regler for anvendelsen af data, især personfølsomme data
 - Forskerne har kendskab til ophavsretslige regler for videnskabelige data
3. Genbrug af data
 - Der sker genbrug af egne data
 - Adgang til andres data: Der er ikke overordnede aftaler om adgang til samlede datamængder og arkiver eller data hos andre forskere – det sker kun på individuelt plan
4. Værktøj
 - Der er et behov for information om programmer, analyser og dataformater
 - Software er kostbar for den enkelte forsker
 - Software er tidskrævende at sætte sig ind i for den enkelte forsker
5. Forskningsmetoder
 - Det er vanskeligt for udenforstående at få indblik i forskernes individuelle analysemetoder og databearbejdning
 - Dette kræver fokus på fag, hvor bibliotekets kompetencer selv er fagligt på højde med forskningen
6. Forskerservice
 - Hjælp er nødvendig og efterspurgt, særlig hjælp mht. punkt 1 og 4 ovenfor
7. Fordele ved dataregistrering og -arkivering
 - Forskerne er bekendt med, at der er fordele ved at dele data.

Endeligt, til brug for analysen af de tre projektspor, har det været nødvendigt, at definere et fælles sæt ledestråde, der tager udgangspunkt i de sekundære hypoteser.

Ledetråde for interviews

1. Tilvejebringelse af data

- a. Fremskaffelse af data eller empiri
 - b. Datamængder
 - c. Genbrug af data (egen og andres)
 - d. Typer af data
2. Databearbejdning
 - a. Hvordan analyseres data eller empiri
 - b. Dataformater og -typer
 - c. Bevaring af informationer om data
3. Software og værktøjer
 - a. IT-værktøjer og software
 - b. Adgang til og køb af IT-udstyr og software
 - c. Får forskeren eller forskningsprojektet support – og af hvem
4. Samarbejde med andre forskere
 - a. Omfang: For eksempel frekvens og mængde – er det essentielt
 - b. Udveksling af data: For eksempel hvilke aftaler
 - c. Hvordan foregår samarbejde
5. Bevaring
 - a. Opbevaring: Sikkerhed – fysisk og elektronisk
 - b. Datamængde
 - c. Adgang: Hvordan og for hvem – kontrol af identitet.

4. Litteraturstudium

Der har været afholdt i alt fire seminarer. På disse seminarer er en del litteratur om primærdata, anvendelse af IT i forskningsprocesser samt udvikling af e-science inden for forskellige videnskabelige områder blevet gennemgået og diskuteret. E-science betegner en kollektiv forskningsproces, hvor IT udgør det element, som binder processens forskellige faser sammen til en helhed, og som binder deltagerne sammen i et virtuelt netværk. I det følgende vil litteraturen blive analyseret med henblik på at bekræfte eller afkræfte de sekundære hypoteser – om muligt.

Inddelingen følger inddelingen af de sekundære hypoteser ovenfor. Der blev under interviewene stillet spørgsmål, som havde fokus på blandt andet de sekundære hypoteser. Der henvises derfor til de allerede nævnte ledetråde under punkterne nedenfor. Desuden henvises i nogle tilfælde til den egentlige interviewguide – denne findes som bilag.

Bevaring

Den sekundære hypotese er at “data forsvinder eller er udsat for utilstrækkelig bevaring”.

Ledetrådene stiller spørgsmål angående opbevaringssikkerhed (fysisk og elektronisk), datamængder og adgange (hvordan, hvem og kontrol af identitet) (ledetråde 5a-5c ovenfor)

Litteraturen⁶ viser, at der er et udtalt behov hos forskerne for langtidsbevaring af data og for deling af data forskerne imellem. De ønsker herudover politikker udviklet vedrørende brugen af et e-science system og standarder for metadata. Men de ønsker også backup eller spejling af data, både af sikkerheds- og adgangsmæssige hensyn.

Regler

De sekundære hypoteser er, at “Forskerne er opmærksomme på datatilsynets og lovgivningens regler samt etiske regler for anvendelsen af data, især personfølsomme data” og at “Forskerne har kendskab til ophavsretslige regler for videnskabelige data”.

Ledetrådene peger på spørgsmål angående problemer med fremskaffelse af primære og sekundære data. Disse kunne for den enkelte forsker omfatte ovenstående sekundære hypoteser.

Her siger litteraturen, at forskerne har brug for – på en simpel måde – at kunne håndtere rettigheder for anvendelse af andres og egne data, men også håndtering af rettigheder til brugen af software på tværs af grupper og landegrænser er ønskeligt. Herudover påpeges det, at der ofte er problemer knyttet til på den ene side at beskytte data og afledte data mod ulovlig tilgang og på den anden side værktøjets pålidelighed⁶.

⁶ Newhouse et al. (2007)

Genbrug af data

De sekundære hypoteser er, at "Der sker genbrug af egne data" og at "Adgang til andres data: Der er ikke overordnede aftaler om adgang til samlede datamængder/arkiver eller data hos andre forskere – det sker kun på individuelt plan".

Ledetrådene peger på spørgsmål angående genbrug af data (egne og andres) (1c) og udveksling af data: For eksempel hvilke aftaler (ledetråd 4b ovenfor).

Litteraturen siger: Forskerne anbefaler, at der skabes politikker omkring udvikling af best practices i forhold til udvikling af præserveringsværktøj og fildeling, fælles standarder for annotering inden for de enkelte forskergrupper, at der dannes definitioner af roller, handlinger og relationer i en virtuel organisation.⁶ Så der er et udtalt behov for at dele data med hinanden og præservere data. Inden for arkæologi er det et behov for at kunne samarbejde og udveksle ideer omkring data. Det er essentielt for et stykke e-Science værktøjs succes ikke at kunne skabe data, men at kunne samarbejde om dem⁷. Forskerne ønsker software, der automatisk kan skabe metadata (annotering og data provenance), nem og kontrolleret fildeling mellem grupper, software til dannelse af virtual organizations, rettighedshåndtering med et simpelt og brugervenligt interface. For at kunne dele data med hinanden, er det nødvendigt at registrere metadata metodisk⁸. Artiklen beskriver en taksonomi for beskrivelse af metadata og hvad man kan bruge metadata til. *"En utilstrækkelig metadatamodel kan føre til manglende evne til at kombinere data fra forskellige kilder, hvilket resulterer i dårlig navigering og dårlig federated søgning i informationen og dermed ringe udnyttelse heraf"*⁹.

⁷ Kilbride (2006)

⁸ Simmhan et al. (2005)

⁹ Hert et al. (2007)

Værktøj

De sekundære hypoteser er, at “Der er et behov for information om programmer, analyser og dataformater”, at “Software er kostbar for den enkelte forsker” og at “Software er tidskrævende at sætte sig ind i for den enkelte forsker”.

Ledetrådene peger på spørgsmål angående software og værktøjer: IT-værktøjer og software, adgang til og køb af IT-udstyr og software, får forskeren eller forskningsprojektet support – og af hvem? (ledetråde 3a-3c ovenfor)

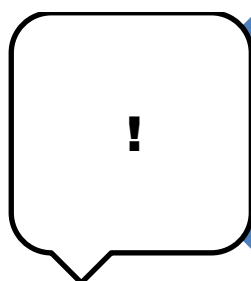
Litteraturen siger: Forskerne ønsker bedre support til brugere, der ønsker at flytte deres lokale applikationer ind i en e-infrastruktur, og til udviklere, der vil optimere deres e-infrastruktur. De ønsker også tilgængeligt selv-hjælps-materiale og tutorials til anvendelse af software – især forskere på områder, der traditionelt ikke har brugt e-science værktøj.⁶ Mere om support vedr. software i afsnit om forskerservice nedenfor.

Forskningsmetoder

De sekundære hypoteser er, at “Det er vanskeligt for udenforstående at få indblik i forskernes individuelle analysemetoder og databearbejdning” og at “Dette kræver fokus på fag, hvor bibliotekets kompetencer selv er fagligt på højde med forskningen”.

Ledetrådene peger på spørgsmål angående forskningsprocessen. I interviewet bliver forskeren bedt om at beskrive processens enkeltdelen og hvilke problemer der er tilknyttet disse. Se interviewguiden – vedlagt som bilag (jf. Bilag C).

Litteraturen siger: Forskerne bruger mange forskellige services inden for de enkelte fag og på forskellige stadier af forskningsprocessen.¹⁰ Litteraturen omtaler ikke denne problematik indgående. Der er dog påpeget et problem af denne art, nemlig at der findes et skel mellem brugere og udviklere af Virtual Research Organizations. Her anbefales det, at “champions” i forskningsmiljøet er med til at drive det frem, og at man involverer brugerne i udviklingen af det. Projektleder skal være et medlem af forskningsmiljøet og ikke en udvikler. Værktøjet skal være simpelt og anvendeligt, og man bør ikke vente på, at det er perfekt. Man bør begynde at anvende det så tidligt som muligt og så løbende udvikle og forbedre det i forhold til de enkelte projekter.



Værktøjet skal være simpelt og anvendeligt, og man bør ikke vente på, at det er perfekt. Man bør begynde at anvende det så tidligt som muligt og så løbende udvikle og forbedre det i forhold til de enkelte projekter.

¹⁰ Voss (2007)

Forskerservice

Den sekundære hypotese er, at "Hjælp er nødvendig og efterspurgt, særlig hjælp mht. bevaring og værktøj".

Ledetrådene peger på spørgsmål angående forskningsprocessens dele og problemer knyttet hertil. Hvis forskerne oplever, at der er brug for support især i forbindelse med bevaring af data og værktøj, vil det fremgå af deres svar vedr. disse aspekter (ledetråde 1 & 4 ovenfor).

Litteraturen siger: Support og undervisning er et stærkt ønske hos forskerne, både til fildeling, annotering og provenance, samt rettighedshåndtering. Der er herudover et stort ønske om bedre information om, hvilke softwareprodukter der er til rådighed på markedet, og hvilke der passer bedst til den enkeltes behov. Dette medfører ofte, at der træffes uinformede valg af software. Der er i høj grad brug for undervisning og undervisningsmaterialer til systemerne. Der er mangel på viden om tilgængelige kurser, som for eksempel software udviklerne eller andre kursusudbydere afholder.⁶ Forskerne ønsker værktøj, der er intuitivt og simpelt at bruge, og det skal findes i deres "comfort zone" – og hvis det fejler, skal det være klart, hvad der går galt, så de rette personer kan kontaktes. Den support, disse brugere efterspørger, er den der er indbygget i et robust stykke software i form af informative fejlmeddelelser. Sikkerhed er også et issue, og forskerne ønsker også undervisning på dette område.¹¹

Fordele ved dataregistrering og -arkivering

Den sekundære hypotese er, at "Forskerne ved, at der er fordele ved at dele data".

Ledetrådene peger på spørgsmål angående samarbejde med andre forskere (ledetråde 4a-4c ovenfor).

Litteraturen siger: Forskerne er faktisk ikke altid villige til at dele deres data, da det indebærer omkostninger i forhold til kvalitetskontrol, skabelse af metadata og langtidsbevaring. Fordelene ved e-science kan for den enkelte forsker være uklare inden for humaniora og samfundsvidenskab. *"Der er en signifikant kontinuitet mellem udviklingen af teknologier og deres anvendelse. Det er vigtigt at forstå de forandringer, den ny teknologi skaber; at de påvirker forskningens sociale organisering eller møder modstand fra denne."*¹²

Der er dog et udtalt behov for og ønske om at dele data med hinanden og samarbejde omkring disse data på andre områder, som nævnt ovenfor⁶, og det er vigtigt, at delingen af data kan foregå på en sikker og overskuelig måde. Især er der behov for værktøjer, der kan hjælpe forskerne i deres daglige arbejde, dvs. som kan udføre job-kørsler (særlige beregninger eller konverteringer af data) og facilitere fildeling. Der behov for værktøjer, som kan administrere læse-, skrive- og

¹¹ Schopf (2005)

¹² Voss (2007)

redigeringsrettigheder – og desuden for værktøjer, som kan visualisere data på måder, som letter forståeligheden.

5. Praksisstudium

Dansk Data Arkiv

Dansk Data Arkiv (DDA) er et nationalt serviceorgan for den samfundsbeskrivende videnskab, det vil sige, for samfundsvidenskab, sundhedsvidenskab og historie, og udgør en del af Statens Arkiver (SA). DDA indhenter, bevarer og udleverer forskningsdata fra de nævnte områder.

Kvantitative forskningsdata i form af spørgeskemaundersøgelser udgør størstedelen af samlingen. DDA's hovedaktiviteter er fokuseret på at hjælpe forskere og studerende med at bruge data og at aflevere data. DDA definerer datamateriale som data og dokumentation, og DDA har både en opgave i forhold til bevaring af data (talmatrix, der udgør substansen i kvantitativt datamateriale) og en opgave i forhold til dokumentation (beskrivelsen af datas tilblivelse og indhold) af data.

Beholdningen består af data og dokumentation fra cirka 2000 undersøgelser indenfor de ovennævnte områder. DDA bevarer og formidler elektroniske data, der er skema- eller interviewbaseret og systematisk indsamlet, og som kan dokumenteres, til brug for sekundære analyser. Både arkivering og udlevering er gratis.

Studier der modtages:

- Tværsnitsstudier
- Registerstudier
- Kombinationsstudier
- Studier baseret på åbne databaser
- Kohorte studier
- Tværnationale studier.

Ved indsendelse af data, udfyldes et afleveringsskema hvor der skal oplyses om

- Titel og ansvarlig
- Beskrivelse af undersøgelsen
- Dataindsamlingsmetode og tidsrum for denne
- Dokumentation af undersøgelsen (findes/medsendes: Spørgeskema, interviewinstruktion, vægtning m.m.)
- Betingelser for adgang til data, offentliggørelse og betingelser for adgang til materiale. Det er den enkelte forsker, der bestemmer reglerne for adgangen til sine data.

Udover at være et arkiv for samfundsbeskrivende forskningsdata har DDA også til opgave at bevare data til fremtidige generationer, at oplyse om muligheder for at genanvende datamaterialer, at rådgive om datadokumentation, og at deltage i internationalt samarbejde med andre databanker.

I forhold til formidling af DDA's materiale, så vil man gerne udvikle undervisningsmateriale til studerende og forbedre søge- og analysemulighederne og derigennem også udbrede kendskabet til materialerne i DDA. Desuden er der også ønske om at få mere kvalitativt materiale ind i DDA. Data er her transskriptioner af optagne interviews. Men hovedvægten ligger som nævnt indtil videre på kvantitative data. DDA arbejder endvidere på at fremme et arkivnetværk for forskere fra blandt andet samfundsvidenskabelige institutter.

DDA påpeger selv, at bearbejdning af data til opbevaring og genbrug er en meget ressourcekrævende proces. Processen er som følger

- Indlæsning af modtagne data
- Formel kontrol
- Omkodninger og dokumentation af data
- Bearbejdning af data og kodebog
- Lagring og udlevering.

Efterspørgslen efter de enkelte datamaterialer bestemmer den rækkefølge, hvori de tages ind til oparbejdning. Der bruges gennemsnitligt cirka 200 timer pr. sæt datamateriale. Det meste af materialet bliver min. brugt en gang, men der er en naturligvis materiale, der er mere "populært" end andet, såsom data fra valg og værdiundersøgelser. Desuden er der stor efterspørgsel efter tværnationale data, som for eksempel Eurobarometer.

Sociologisk Databank (SD), Sociologisk Institut

Undersøgelsen er dels foretaget ved at gennemse databankens hjemmeside¹³ og ved at interviewe den ansatte (studerende på Sociologisk Institut, 10 timer ugentligt). Data i banken er videnskabelige data, der er tilgængelige i forvejen – fra for eksempel DDA og Det Nationale Forskningscenter for Velfærd (SFI) – eller som banken har fået tilladelse til at bearbejde og offentliggøre.

Fra databankens hjemmeside:

"Fra Sociologisk Databank kan du downloade forarbejdede kvantitative data til din forskning, undervisning eller dine opgaver inden for en række forskningsområder. Databanken gør det nemmere for dig at udføre statistiske analyser af sociale, kulturelle, økonomiske og politiske forhold - også på tværs af lande og over tid. Du kan anvende databankens materiale til kvantitative undersøgelser eller som supplement til kvalitative problemstillinger."

Databanken er først og fremmest beregnet på studerende ved Sociologisk Institut, Den er til brug ved studiet og skal bidrage til at udbygge kendskab til og viden om kvantitative data og brug af disse.

¹³ <http://databank.soc.ku.dk>

Data er opdelt i

- Forskningsemner
- Fulldataoversigt
- Eksamensdata
- Undervisningsdata.

Banken er tænkt som en understøttelse af den kvantitative forskning, som er et af instituttets satsningsområder. Forskere kan indlevere deres data fra deres egen forskning, men banken indeholder overvejende data fra for eksempel DDA og har links til nationale og internationale databaser. Altså data der i forvejen er behandlet og gennemarbejdet. Det vil sige, at alle data er offentligt tilgængelige, eller at Sociologisk Institut har fået tilladelse til at bruge de i banken tilgængelige data. Så der brydes ingen regler ved at bruge data fra banken.

Hjemmesiden indeholder desuden en udførlig vejledning i, hvordan man finder data, downloader dem og åbner dem etc. Desuden ydes der hjælp til ansatte og studerende i forbindelse med banken, såsom mindre statistiske analyser, SPSS og hjælp til undervisere med at konstruere datasæt til brug i undervisningen. Desuden er der vejledning til, hvordan en problemstilling kan løses ved hjælp af datasæt.

Banken kan bedst beskrives som en formidlingsdatabank, der formidler information om data og links til statslige arkiver m.m. som de tidligere nævnte – og er baseret overvejende på allerede gennemarbejdede datamaterialer.

6. Interviews

Der er i alt foretaget ni interviews. De seks er optaget på bånd og diktafon, og derefter udskrevet. De resterende tre foreligger kun som notater fra interviewereren – som følge af forskellige tekniske problemer. I alt interviewes ti personer, da der i et enkelt tilfælde interviewes to personer.

De interviewede er alle forskere fra Det Samfundsvidenskabelige Fakultet ved KU og fordeler sig således på fag – fire fra Institut for Psykologi, to fra Institut for Statskundskab, én fra henholdsvis Økonomisk Institut, Sociologisk Institut og Institut for Antropologi (jf. Bilag D).

De interviewede fordeler sig således på ansættelseskategorier – to professorer, syv lektorer og én ph.d.-studerende. I det tilfælde, hvor to personer interviewes, er der tale om en professor og én ph.d.-studerende.

Alle interviews resumeres således: Først beskrives hvad forskerne gør, derpå gengives hvad forskerne mener, ønsker eller har behov for.

Interview 1 (psykologi)

Forsker 1 indsamler og arbejder med data fra spørgeskemaer og interviews (både mere strukturerede og mere kvalitative). Spørgeskemaer er på papir – data inddateres efterfølgende. Genbruger ikke data. Opbevaring af elektroniske data sker på fællesdrevet. Mængden af data giver ikke problemer, men data er personfølsomme og skal destrueres efter en vis tid. Data behandles med statistikprogrammet SPSS. Der er adgang til SPSS på instituttet, og der undervises i programmet, men andre værktøjer må de enkelte forskere selv opspore og lære. Samarbejder primært med kolleger på instituttet, og data sendes med mail. Deltager i virtuelt netværk, men ikke aktivt.

Forsker 1 har behov for mere support til opdagelse, valg og brug af nye programmer – for eksempel webbaserede spørgeskemaer. Mener, at forskere på instituttet har meget forskellige behov, hvad angår redskaber til dataindsamling og behandling. Mener desuden, at opbevaring af data et særligt sted kunne være nyttig, forudsat at der er adgangsregulering og procedurer for anonymisering, destruktion af data m.m.

Interview 2 (statskundskab)

Forsker 2 indsamler og arbejder med data fra kvalitative spørgeskemaer og interviews – og benytter også allerede foreliggende statistiske data. Indsamlede data fra spørgeskemaer og interviews opbevares (sidstnævnte som lydfiler) dels på egen PC og dels på fællesdrevet – mens de statistiske data er placeret hos Danmarks Statistik. Mængden af data, som ligger lokalt, giver ikke problemer – men data er personfølsomme. Data behandles med programmet Nvivo og med statistikprogrammet SPSS. Hvad angår lokal support til opdagelse, valg og brug af programmer, så hjælper forskeren sig selv, går på kursus, får hjælp fra studerende eller fra kolleger uden for KU – men programmerne har givet en del besvær. Deltager i samarbejde nationalt og internationalt,

men udveksler og deler ikke rådata med andre. Genbruger allerede eksisterende data fra Dansk Data Arkiv (DDA) i forbindelse med undervisning.

Forsker 2 har behov for mere support til opdagelse, valg og brug af programmer.

Interview 3 (antropologi)

Forsker 3 indsamler og arbejder med data fra interviews og surveys – og benytter også i en vis udstrækning allerede foreliggende statistiske data (hvis de findes og er troværdige). Producerer og anvender desuden billeder og film, samt data fra deltagerobservationer. Indsamlede data opbevares på egen PC og på andre projektdeltageres PC'er (for eksempel transskriberede interviews). En slags stikordsregister udarbejdes, således at der kan søges på tværs af forskellige interviews. Mængden af data giver ikke problemer. Data behandles med programmet AskSam og tekstbehandling – statistikprogrammer som SPSS er ikke aktuelle. Også bibliografiprogrammer og Endnote har interesse. Hvad angår opdagelse, valg og brug af programmer, så hjælper de enkelte forskere sig selv eller hinanden, og det fungerer nogenlunde – forskerne bruger normalt de redskaber, de plejer at bruge. Deltager i samarbejde nationalt og internationalt. Deler normalt kun behandlede data med andre – og data fra deltagerobservationer er både følsomme og vanskelige at dele.

Forsker 3 ønsker, at billeder og evt. også film indskannes og opbevares et særligt sted – og mener, at disse vil være nyttige for både forskning og undervisning - men adgangs begrænsning er nødvendig, og et stikordsregister ville være nyttigt. Mener desuden, at opbevaring af andre data kunne være nyttig, fordi opbevaring på egen PC er risikabel. Mener endvidere, at ordentlig og sikker backup er vigtigere end nye IT-værktøjer – og at problemet med teknikere er, at de ikke forsker. Mener endelig, at langtidsopbevaring er vigtig, det vil sige, opbevaring på tværs af skiftende teknologier og programmer – og at forskning på basis af andres rådata måske vil spille en voksende rolle i fremtiden.

Interview 4 (statskundskab)

Forsker 4 arbejder med data i form af dokumenter – det vil sige, genbruger og behandler allerede eksisterende data. Dokumenter fremskaffes gennem litteratursøgning og egen eller andres viden. Nogle få dokumenter er elektroniske, hovedparten er trykt. Opbevaring af elektroniske data sker på egen PC og på fællesdrevet. Bruger ikke (speciel) software. Deltager i samarbejde nationalt og internationalt. Udveksling og deling af rådata finder ikke sted – men behandlede data sendes med mail til samarbejdspartnere. Deltager ikke i virtuelle netværk.

Forsker 4 har ingen særlige behov for (mere speciel) software og support.

Interview 5 (økonomi)

Forsker 5 arbejder med data fra Danmarks Statistik – det vil sige, genbruger og behandler allerede eksisterende data. Opbevaring af behandlede data sker på egen PC og på fællesdrevet. Mængden af data giver ikke problemer, men data er personfølsomme. Data behandles med forskellige

statistikprogrammer – og instituttet indkøber og stiller til rådighed de programmer, som forskerne har brug for. Deltager i samarbejde nationalt og internationalt. Udveksling og deling af rådata finder ikke sted – på grund af den karakter data har, og på grund af de krav Danmarks Statistik stiller – men behandlede data sendes med mail til samarbejdspartnere. Deltager ikke i virtuelle netværk.

Forsker 5 ønsker en database med datasæt, som de studerende kan bruge i forbindelse med undervisning.

Interview 6 (psykologi)

Forsker 6 indsamler og arbejder med data fra samtaler med klienter i klinik (journaler), samt data fra interviews og webbaserede spørgeskemaer. Data fra klienter suppleres med data fra terapeuter eller forskere. Data fra spørgeskemaer indsamles i samarbejde med Det Nationale Forskningscenter for Velfærd (SFI). Data er personfølsomme, men anonymiseres. De elektroniske data opbevares på PC'er og i en databank på Institut for Folkesundhed ved Syddansk Universitet. I forbindelse med behandlingen af data anvendes regneark og statistikprogrammet SPSS. Lokal support i forbindelse med opdagelse, valg og brug af programmer præges af individuelle løsninger og personlige kontakter. Deltager i nationalt og internationalt samarbejde. Udveksling og deling af data sker gennem mails til samarbejdspartnere.

Forsker 6 har behov for mere support til opdagelse, valg og brug af programmer – herunder ikke mindst vedr. webbaserede spørgeskemaer. Mener, at forskere på instituttet har meget forskellige behov, hvad angår redskaber til dataindsamling og behandling. Mener desuden, at opbevaring af data et sikkert sted kunne være nyttig, forudsat at der er adgangsbegrænsninger. Mener endelig, at langtidsopbevaring kunne være nyttig, det vil sige, opbevaring på tværs af skiftende teknologier og programmer – men at nogle data bliver mindre interessante over tid.

Interview 7 (psykologi)

Forsker 7A og 7B indsamler og arbejder med forskellige typer af data – data fra interviews (optagelser), observationer (nedskrevne), surveys (internetbaserede spørgeskemaer) og tests. Arbejder desuden på, at test- og surveydata fra psykologistuderende (gennem en længere årrække) bliver gjort tilgængelige, det vil sige, kan genbruges i forbindelse med især undervisning, men også forskning. Data er personfølsomme. Nogle data opbevares på en særlig server, nogle data opbevares på egen PC og andres PC'er – og data fra studentertests opbevares hos Dansk Psykologisk Forlag. Bearbejdede data ligger hos de personer, som bearbejder dem – og forskere opbevarer ikke blot rådata, men også forskellige faser i bearbejdningen af rådata. Data indsamles og bearbejdes med blandt andet SurveyXact (spørgeskemaprogram), statistikprogrammet SPSS og Excel. Også referencehåndteringsprogrammer har interesse. Hvad angår lokal support til opdagelse, valg og brug af programmer, så spiller individuelle løsninger og personlige kontakter den afgørende rolle, og processen præges af tilfældighed og uigennemsigthed. Deltager i nationalt og internationalt samarbejde. Udveksling og deling af data – samt kommunikation vedr.

interviewguides og spørgeskemaer – sker især gennem mails. Virtuelle netværk benyttes ikke – men Google-grupper benyttes i et vist omfang.

Forsker 7A og 7B har behov for mere effektiv og bedre organiseret support vedr. opdagelse, valg og brug af programmer. Mener, at opbevaring af data et særligt sted kunne være nyttig – forudsat at der er begrænset adgang og beskyttelse mod hacking. Mener desuden, at langtidsopbevaring af data kunne være nyttig, det vil sige, opbevaring på tværs af skiftende teknologier og programmer – men at der skal være særlige procedurer for, hvor længe data opbevares.

Interview 8 (psykologi)

Forsker 8 indsamler og arbejder med data fra forsøgspersoner, det vil sige, fra spørgeskemaer og tests - og data fra datafiler. De førstnævnte starter på papir og bliver senere elektroniske, mens de sidstnævnte fra starten findes i en database. Data er personfølsomme. Data opbevares forskellige steder – på egen PC eller andres PC'er og i en hoveddatabase på Rigshospitalet. Data bearbejdes især ved hjælp af regneark og statistikprogrammet SPSS. Hvad angår lokal support til opdagelse, valg og brug af programmer, så spiller faglige, individuelle og personlige relationer og kontakter den afgørende rolle. Deltager i nationalt og internationalt samarbejde. Udveksling og deling af data sker især gennem mails. Indgår ikke i virtuelle netværk.

Forsker 8 har behov for mere support i forbindelse med opdagelse, valg og brug af programmer, men kun hvis supporten har et indgående kendskab til psykologisk forskning. Mener, at opbevaring af data i lang tid er vigtig af flere forskellige grunde – men hvor længe afhænger af, hvilken type data der er tale om. Mener også, at opbevaring af data et særligt sted kunne være nyttig, fordi opbevaring på PC er sårbar – men der skal være forskellige former for adgangsbegrænsninger, og det skal fungere både effektivt og ikke-bureaukratisk.

Interview 9 (sociologi)

Forsker 9 indsamler og arbejder med kvantitative data fra spørgeskemaer og kvalitative data fra interviews. De førstnævnte er elektroniske, de sidstnævnte findes som lydfiler og udskrifter af disse. Interviewdata kan være personfølsomme – men ofte vil den interviewede give tilladelse til publicering. Spørgeskemadata opbevares forskellige steder – herunder databanker og -arkiver i udlandet og Dansk Data Arkiv (DDA). Opbevarer altid kvantitative data i DDA – og genbruger også andres data via DDA. Kvalitative og lokalt opbevarede kvantitative data opbevares på fællesdrev. Kvantitative data bearbejdes ved hjælp af forskellige statistikprogrammer, herunder SPSS. Kvalitative data bearbejdes uden brug af avanceret teknik. Hvad angår lokal support til opdagelse, valg og brug af programmer, så spiller personlige kontakter den afgørende rolle, men instituttet har en overordnet politik mht. indkøb af statistikprogrammer. Deltager i nationalt og internationalt samarbejde. Udveksling og deling af data (samt udkast til tekster m.m.) sker blandt andet gennem et grupperum i punkt.ku – og gennem direkte download fra eksterne institutioner og arkiver. Benytter ikke internationale virtuelle netværk.

Forsker 9 ønsker en databank med kvalitative data især til brug for studerende – som supplement til Sociologisk Databank (med kvantitative data) – men er skeptisk mht. om bankerne bør ligge hos biblioteker. Mener, at mere support i forbindelse med opdagelse, valg og brug af programmer kunne være nyttig – men at der bør være en betydelig valgfrihed, hvad angår programmer, og at supporten ikke må være bureaukratisk og centralistisk. Mener desuden, at valg af programmer er tæt forbundet med forskningsprocessen, og er skeptisk mht. om biblioteker har tilstrækkelige kompetencer på området. Ønsker måske snarere mere support fra DDA, det vil sige, at DDA bliver i stand til at levere mere end rådata.

Opsummering på tværs

De elektroniske rådata, som indsamles og opbevares lokalt af forskere, er især data fra spørgeskemaer, tests, interviews og observationer – det vil sige, der er tale om tekster, regneark, skemaer m.m., men også om lydfiler, billeder og film m.m. Hovedparten af data har en kvalitativ karakter – men det gælder formentlig ikke alle. De opbevares på egen PC, andres PC'er eller et fællesdrev – og i mange tilfælde tilsyneladende uden særlig effektiv eller velorganiseret back up. En del data er personfølsomme – det vil sige, hvis de skal kunne benyttes af andre, skal de anonymiseres, eller der skal indhentes tilladelser dertil.

Nogle data indsamles og/eller opbevares af eller i samarbejde med eksterne institutioner som for eksempel Dansk Psykologisk Forlag, Danmarks Statistik (og tilsvarende i andre lande), Rigshospitalet, Det Nationale Forskningscenter for Velfærd (SFI), Dansk Data Arkiv (DDA) og Syddansk Universitet – eller stilles til rådighed af disse institutioner, og er evt. underlagt begrænsninger fastlagt af disse institutioner.

Normalt er den lokalt opbevarede datamængde ikke et problem – måske fordi data indsamles og/eller opbevares af eksterne institutioner i situationer, hvor der er tale om store datamængder.

Data, som er interessante for forskere, er dels data som kan anvendes i forbindelse med egen eller andres forskning, dels data som kan anvendes i forbindelse med egen eller andres undervisning, og dels data som kan anvendes af studerende. De forskellige typer data er ikke identiske, men kan være det i nogle tilfælde – og hvor der er tale om data til brug for undervisning eller studerende, skal disse formentlig kunne deles med andre i større omfang og på en anden måde, end hvor der er tale om data til brug for forskning.

Den lokale behandling af data sker gennem anvendelse af forskellige programmer – og support i forbindelse med disse programmer er enten i høj grad overladt til forskerne selv eller præget af lidt tilfældige personlige relationer, og hele processen forekommer uigennemskuelig og uoverkommelig for flere forskere. Der synes at være et behov for yderligere support – men i mange situationer skal denne formentlig ikke have en teknisk, men en faglig-teknisk karakter, og forskellige forskere vil formentlig have ret forskellige behov, hvad angår programmer og support.

Nogle forskere udveksler og deler kun bearbejdede data, mens andre også udveksler og deler rådata. Udvekslingen sker normalt som mails – hvor der ikke er tale om direkte download fra en

central institution (databank, dataarkiv m.m.). Virtuelle netværk og Google-grupper m.m. benyttes ikke eller kun lidt.

Flere forskere ser positivt på at opbevare data et særligt sted og over længere tid – således at disse data kan genbruges af dem selv eller andre i forbindelse med ny forskning, kontrol af forskning, dokumentation af forskning, metaanalyser, studieopgaver – og i forbindelse med egen eller andres undervisning. Men der lægges vægt på, at der skal være begrænset adgang til data, at data skal være tilgængelige på tværs af skiftende teknologier og programmer, at der skal være beskyttelse mod hacking og procedurer for, hvor længe bestemte data opbevares, og at opbevaringen skal være både effektiv og ikke-bureaukratisk.

Relation til primære og sekundære hypoteser

Interviewresultaterne synes at understøtte nogle af de primære hypoteser, nemlig at en bedre og mere velorganiseret opbevaring af data kan bidrage til øget effektivitet i forskningsprocessen, øget synlighed af aktuel og tidligere forskning og øget genbrug af empiriske data. De synes i mindre grad at understøtte de andre primære hypoteser – det er ikke let at se klare tegn på nye forskningsformer og tværvideenskabelig forskning, nye strukturer og teknologier, nye klart definerede arbejdsgange (workflows) – især på tværs af de enkelte forskere/forskningsprojekter.

Interviewresultaterne synes at understøtte nogle af de sekundære hypoteser, nemlig at der er problemer mht. bevaring af data, at forskerne er opmærksomme på regler for anvendelsen af (personfølsomme) data, at genbrug af egne data er nyttig, at adgang til andres data ikke sker på en systematisk og velorganiseret måde (bortset fra situationer hvor adgangen sker gennem eksterne institutioner eller arkiver), at der er fordele ved at dele data, at der er et behov for information om programmer m.m., at software er dyr og tidskrævende – og at der er behov for øget forskerservice på nogle af disse områder. De synes i mindre grad at understøtte de andre sekundære hypoteser – det gælder forskernes kendskab til ophavsretslige regler, og tesen om en faglig match mellem bibliotek og forskning inden for visse fag.

7. Konklusion og anbefaling

De interviewede forskere er ret forskellige hvad angår brug af IT i forskningsprocessen (udover tekstbehandling m.m.) – både hvad angår typen og mængden af hjælpemidler – men de synes at have det til fælles, at IT netop anvendes som hjælpemiddel og ikke indgår som en integreret og konstituerende del af forskningsprocessen. Egentlig e-science synes der derfor (endnu) ikke at være tale om.

Flere af de interviewede forskere har brug for yderligere support til opdagelse, valg og brug af programmer til indsamling bearbejdning af primærdata – men den ønskede support skal have en fagligt-teknisk karakter og forskernes behov for programmer og support kan være meget forskellige.

Flere af de interviewede forskere har brug for opbevaring af data et sikkert sted og over længere tid – af en række forskellige årsager og både af hensyn til egen forskning, undervisning og studieopgaver. Set i relation til universitetsbibliotekets kompetencer forekommer det mest oplagt at sætte et konkret pilotprojekt i værk på dette område.

DDA opbevarer for øjeblikket kvantitative data – det forekommer derfor mest hensigtsmæssigt for universitetsbiblioteket primært at satse på opbevaring af kvalitative data – samt evt. på kvantitative data som af en eller anden årsag ikke er interessante for DDA eller som primært er interessante i forbindelse med undervisning eller studieopgaver. En databank til brug for studerende findes på Sociologisk Institut, men der synes at være behov for noget tilsvarende på andre samfundsvidenskabelige institutter.

Det foreslås derfor at der igangsættes et konkret opfølgende pilotprojekt, og at der foretages konkrete forsøg med opbevaring og formidling af data, eksempelvis:

- Billeder (og måske også lokalt producerede film) fra Institut for Antropologi (jf. interview 3)
- Data fra Institut for Psykologi (jf. interview 7 og interview 1).

Hvis dette ikke kan lade sig gøre eller skønnes hensigtsmæssigt, foreslås det, at der i stedet gøres forsøg med opbevaring og formidling af følgende data i forbindelse med et evt. opfølgende projektet:

- Data fra Institut for Psykologi (jf. interview 7 eller interview 1)
- Studenterdata til brug for undervisning af psykologistuderende (jf. interview 7) – hvis det kan lade sig gøre – disse opbevares pt. helt eller delvist af Dansk Psykologisk Forlag.
- Kvantitative data til brug for undervisning af økonomistuderende (jf. interview 5) – hvis det kan lade sig gøre – disse opbevares pt. helt eller delvist af Danmarks Statistik.
- Kvalitative data til brug for sociologistuderende (jf. interview 9).

Fokus bør primært være bevaring og den personlige adgang til data, med mulighed for deling og udvikling af fremtidige overlay services: Et forskningens "myspace".

Bilag A: Deltagere i projektet

Projektstyregruppe:

- Tina Pipa (KUBIS-SAMF)
- Birte Christensen-Dalsgaard (KB-DIS)
- Michael Cotta-Schønberg (KUBIS)
- Bertil Dorch (KUBIS): Formand

Projektarbejdsgruppe og roller:

- Bertil Dorch (KUBIS): Projektleder
- Jens Ludvigsen (KUBIS-SAMF): Praksisstudium (Dansk Data Arkiv & Sociologisk Databank)
- Simone Schipp von Branitz Nielsen (KUBIS-SAMF/VIVIKOM): Interviews og diverse
- Sten Barfort (KUBIS-SAMF/VIVIKOM): Opsummering af interviews, samlet rapport
- Anders Conrad (KB-DIS): Anker for DIS
- Sigfrid Lundberg (KB-DIS): Litteraturstudium, tovholder for kravspecifikation
- Gitte Petersen (KB-DIS): Interview-guide
- Katja Guldbæk (KB-DIS): Tovholder for Interview-guide
- Birgit N. Henriksen (KB): Kravspecifikation
- Tina Pipa (KUBIS-SAMF): Tovholder for KU-kontakt
- Asger V. Larsen (KUBIS-VIVIKOM): Tovholder for litteraturstudium
- Aiysha Abu-Laban (KUBIS-SAMF): Studentermedhjælp, transskription af interviews
- Christina Vega Pedersen (KUBIS-SAMF): Studentermedhjælp, transskription af interviews

Bilag B: Litteraturliste

- Hert, C. A. (2007). Investigating and Modeling Metadata Use to Support Information Architecture Development in the Statistical Knowledge Network. *Journal of the American Society for Information Science and Technology* . <http://www3.interscience.wiley.com/cgi-bin/fulltext/114265128/PDFSTART>. 30/10 2009
- Kilbride, W. (2006, July). Grand challenges: Grand opportunities? archaeology, the historic environment sector and the E-science programme. *Glasgow Museums* . <http://ahds.ac.uk/e-science/documents/Archaeology-grand-challenges.pdf>. 30/10 2009
- Newhouse, S. (2007, April). Study of User Priorities for e-Infrastructure for e-Research (SUPER). *UK e-Science Technical Report Series* . http://www.nesc.ac.uk/technical_papers/UKeS-2007-01.pdf. 30/10 2009
- Schopf, J. M. (2005). Real Users - Real requirements. *UK eScience All Hands Meeting*. <http://www.allhands.org.uk/2005/proceedings/papers/429.pdf>. 30/10 2009
- Simmhan, Y. L. (2005). A Survey of Data Provenance in e-Science. *SIGMOD Record* , pp. 31-36. <http://www.sigmod.org/sigmod/record/issues/0509/p31-special-sw-section-5.pdf>. 30/10 2009
- Voss, A. (2007). e-Research Infrastructure Development and Community Engagement. <http://www.allhands.org.uk/2007/proceedings/papers/866.pdf>. 30-10 2009.

Bilag C: Interviewguide

Skema 1: Indledning

Informant (navn):

Område (fag):

Ansættelsessted:

Tid	Hovedpunkter	Stikord	Information/spørgsmål
	<p>Præsentation & formål</p>	<p>Om mig og interviewet</p> <ul style="list-style-type: none"> Hvem er jeg? Hvor dette interview? Interview-fokus Hvorfor netop dig? Hvem andre? Hvad sker med info fra dit interview? Rapport færdig 	<ul style="list-style-type: none"> Interviewers navn og stilling Kortlægning af hvad forskerne gør nu og hvad de kunne ønske sig af services, funktionalitet og muligheder i fremtiden. Forskernes praksis for opsamling, bearbejdning og behandling af deres data. Din personlige erfaring med dataanvendelse, samt din viden om forskningsområdet i øvrigt. 2 personer inden for hvert samf. Fag Udskrives og analyseres samlet med de øvrige interviews. Samles i intern rapport. Bliver ikke offentliggjort uden for biblioteksledelsen. Ultimo 2009 Du vil modtage et eksemplar af rapporten.
	<p>Interviewets rammer</p>	<p>Tidsramme, sikring af anonymitet, tilladelse til at optage på bånd.</p> <ul style="list-style-type: none"> Rollefordeling Form 	<ul style="list-style-type: none"> Interview varer 1 time +/- Må vi optage på bånd til brug i analysearbejdet? Anonymisering inden videreformidling. Forsyn dig med kaffe, kage, vand, frugt..... Du er eksperten. Vi vil lære noget af dig, din erfaring og din viden om hvad der sker på området. Formen er samtale-agtig. Du fortæller og jeg spørger undervejs. En række spørgsmål (temaere), men ingen fast rækkefølge - du taler bare Vi er ude efter hvad og hvordan du gør med data – vi dømmes/vurderer ikke hvad du gør (der findes ikke forkerte svar).
	<p>Din egen forskning</p>	<p>Fortæl mig om dit forskningsprojekt</p>	
	<p>Evt. bemærkninger:</p>		<ul style="list-style-type: none"> Her kan forskeren "tale frit fra leveren"

Skema 2: Data-indsamling og bearbejdning

Tid	Hovedpunkter	Stikord	Information/spørgsmål
	<p>Tilvejebringelsen af data - dig/andre</p> <p>Evt. bemærkninger:</p> <p>Databearbejdning</p> <p>Evt. bemærkninger:</p>	<p>Fortæl mere om den data, der indgår i din forskning</p> <ul style="list-style-type: none"> • <i>Fremskaffelse</i> • <i>Opbevaring/lagring</i> • <i>Datamængde</i> • <i>Genbrug</i> • <i>Andre forskere</i> <p>Hvordan bearbejder du efterfølgende dine data?</p> <ul style="list-style-type: none"> • <i>Overblik</i> • <i>Indeksering</i> • <i>Formater</i> • <i>Bevaring af info</i> <p>Hvordan bearbejdes data typisk inden for dit forskningsområde?</p>	<ul style="list-style-type: none"> • Hvilken type data (tal, tekst, interviews, billeder)? • Er data indsamlet specifikt til formålet (primære data)? • Hvordan fremskaffede du data (metode)? • Hvor "gjorde" du af den data, du fremskaffede? (lokal pc, server) • Hvor store (ca) datamængder genererer din forskning? (MB, anden målestok) • Er denne datamængde typisk for dig? For andre forskere inden for dit område? • Stiller datamængden særlige krav til computerkraft? (håndtering, filoverførsel m.m.) • Har du genbrugt egne data fra tidligere forskningsprojekter? • Har du genbrugt andre forskers data indsamlet til andre formål? • Har du brugt data fra eksterne leverandører? (ex. Dansk Data Arkiv, Dansk Statistik) • Hvordan holder du dig opdateret med, hvad der findes af data inden for dit forskningsområde? • Er den data, du har anvendt, typisk for området? • Har du kolleger, der adskiller sig mærkbart fra dig mht. ovenstående? – hvordan? <ul style="list-style-type: none"> • Bearbejder du selv din data? • Hvordan gør du data klar til analyse? • Hvordan får du overblik over din data? • Indekserer/opmærker/annoterer/fuldtekstindekserer du din data? • Indekserer du efter egne eller fælles standarder? • Hvor mange forskellige formater, har dine data været konverteret til? (lyd, tekst m.m.) • Hvordan sammenkæder du de forskellige skridt (formater) i databearbejdningen? • Hvordan bevarer du info om indsamlingsmetode? • Hvordan sikre du info om provenance? <ul style="list-style-type: none"> • Er omtalte databearbejdningsproces typisk inden for dit forskningsområde? • Er der udviklet standard-metoder til håndtering af data inden for dit forskningsområde? • Er der nogle af dine kolleger, der gør ovenstående på helt andre måder?

Skema 3: Om brug af IT-baserede værktøjer

Tid	Hovedpunkter	Stikord	Information/spørgsmål
	<p>Om softwarebrug - dig/andre</p> <p>Evt. bemærkninger:</p>	<p>Fortæl om de IT-baserede værktøjer, du bruger til behandling og analyse af forskningsdata</p> <ul style="list-style-type: none"> Hvilke IT-værktøjer? Hvad bruger du dem til? Kan de tilpasses? Adgang? Finansiering? Viden om nye værktøjer? Erfaringer (gode/dårlige) Nok support? OK sikkerhed? <p>Fortæl generelt om den rolle IT-værktøjer spiller for dit forskningsområde</p> <ul style="list-style-type: none"> Strategi? Accepteret? Generelle holdning? Største udfordring? Største forhindring? <p>Evt. bemærkninger:</p>	<ul style="list-style-type: none"> Bruger du PC eller MAC til din datahåndtering? Hvilke It-baserede værktøjer anvender du? Automatisk annotering eller indeksering? Value - adding? Finde mønstre, tendenser, afvigelser i større datamængder? Skabe overblik over større datamængder? (data mining) Modellering? Simulering? Visualisering? Statistisk bearbejdning? Sammenligning af forskellige data-sæt? Udveksling af større filer? Projektstyring? Kan værktøjerne tilpasses/videreudvikles efter behov, eller er det generelle værktøjer? Er værktøjerne onlinebaserede eller lokale applikationer? Hvordan finansieres adgangen til værktøjerne? Hvordan holder du dig orienteret om nye værktøjer på markedet, som kan understøtte din forskning! Hvordan hørte du første gang om IT-værktøjerne? Hvad er dine erfaringer med værktøjerne? Hvad kan forskere med disse, som de ikke kan uden? Er der mangler ved værktøjerne? Hvad? Har du oplevet problemer? Hvad har du gjort? Er det et problem at få support til værktøjerne? Har du oplevet dit IT-værktøj er blevet forældet? Føler du det sikkert at bruge IT-værktøjerne? Indgår IT-værktøjer i din institutions strategi? Hvilke IT-værktøjer er mest anvendte og anerkendte inden for forskningsområdet? Synes forskerne inden for dit forskningsområde at være klar til at udnytte IT-værktøjer? Hvad er den generelle holdning blandt forskningsområdets forskere til brugen af IT-værktøjer? Hvad er den største udfordring i forhold til brug af IT-værktøjer inden for dit forskningsområde? Hvad er de største forhindringer for udbredelsen af IT-værktøjer?

Skema 4: Samarbejde forskere imellem, samt bevaring af data

Tid	Hovedpunkter	Stikord	Information/spørgsmål
	<p>Samarbejde med andre</p> <p>Evt. bemærkninger:</p> <p>Bevaring af data – dig/andre</p>	<p>Fortæl om dit samarbejde med andre forskere – nationalt/internationalt</p> <ul style="list-style-type: none"> • <i>Med hvem?</i> • <i>Hvor udbredt?</i> • <i>Om hvad?</i> • <i>Udveksling af data, resultater og erfaringer?</i> • <i>Hvordan samarbejder I?</i> • <i>Forskningsnetværk?/samarbejde med kollegaer eller folk med samme interesse/forskningsområde</i> • <i>OK samarbejdsrammer?</i> • <i>Virtuelle fællesskaber?/erfaringsudveksler du med kollegaer via internettet?</i> • <i>Erfaringer med virtuelle fællesskaber</i> <p>Hvad du gør ved data efter endt forskningsprojekt?</p> <ul style="list-style-type: none"> • <i>Opbevaring</i> • <i>Datamængde</i> • <i>Tilgængelighed</i> • <i>Utilgængelighed</i> • <i>Langtidsbevaring</i> • <i>Hvad gør kollegerne?</i> 	<ul style="list-style-type: none"> • Hvilke andre forskere samarbejder du med? Fra andre institutioner og/eller andre lande? • Hvor udbredt er samarbejde på tværs af faglige grupper og professionelle fora inden for dit forskningsområde? • Om hvad samarbejder I? • Diskuterer I jeres forskningsresultater? - Hvordan? • Udveksler I data? - Hvordan? • Udveksler I erfaringer med IT-værktøjer? - Hvordan? • Samarbejder I om IT-værktøjer? • Er du med i en eller anden form for forskningsnetværk eller samarbejdsorganisation? • Hvad bibringer det dig at være med i disse? • Hvilke forskningsnetværk eller samarbejdsorganisationer er mest udbredt inden for dit område? • Er samarbejdsrammerne tilfredsstillende (teknisk/organisatorisk)? Hvad kan optimeres? • Kender du virtuelle forskningsfællesskaber inden for dit forskningsområde? • Er du med i et virtuelt fællesskab? Hvilket? Hvorfor? • Er de forskningsspecifikke/af mere generel karakter? • Hvad er dine erfaringer med virtuelle fællesskaber? - nemt? svært? sikkert? Uoverskueligt? • Er der noget du savner i den forbindelse? • Hvad er de største problemer med at samarbejde? • Er der noget, du kunne tænke dig anderledes? • Hvad gør du med data efter dit forskningsprojekt? • Hvor/hvordan opbevarer du data? • Hvilke datamængder er der typisk tale om? • Giver mængden problemer i forhold til PC størrelse? • Gør du data tilgængelig for andre

Bilag D: Informanter

Informanter fra Københavns Universitet interviewet i forbindelse projektet.

- Lektorer: 7
- Professorer: 2

- Institut for Antropologi: 1
- Institut for Statskundskab: 2
- Institut for Psykologi: 4
- Sociologisk Institut: 1
- Økonomisk Institut: 1