

Karen Lillo

Deling av forskningsdata – forskerens perspektiv

**En intervjustudie med forskere ved norske
forskningsinstitusjoner**

**Masteroppgave 2022
Master i bibliotek- og informasjonsvitenskap**

OsloMet – Storbyuniversitet, Institutt for Arkiv– bibliotek– og informasjonsfag

Sammendrag

Dette masterprosjektet har til hensikt å undersøke hvilke holdninger og erfaringer forskere har med deling av forskningsdata. Jeg var interessert i å få vite mer om hvordan slik datadeling fungerer i praksis, og om det er noen spesifikke faktorer som har innvirkning på forskeres datadeling.

Det oppfordres til deling av forskningsdata gjennom EUs rammeprogram for forskning og for forskning finansiert gjennom Norges forskningsråd. Jeg ønsket å undersøke hvordan forskere responderer på disse oppfordringene, og de utfordringene de eventuelt møter når og hvis de gjør forskningsdata tilgjengelig for andre. Ved å intervju forskere om konkrete utfordringer de har møtt og erfaringer de har gjort seg, håper jeg å bidra til mer innsikt i kompleksiteten rundt deling av forskningsdata.

Jeg har intervjuet 12 forskere, der åtte av forskerne har naturvitenskapelig bakgrunn og fire forskere har bakgrunn fra samfunnsvitenskap/humaniora. Forskerne var i hovedsak positive til å dele data utenfor sin egen forskningsgruppe. Selv om forskerne fra samfunnsvitenskap pekte på at det ligger noen begrensninger i å dele datamaterialet fra deres forskning. Mange av informantene delte oppfatningen om at data som er finansiert med offentlige midler er et kollektivt gode. Men for at forskerne skal være villige til å dele data er det også visse betingelser som må være på plass, i tillegg hadde de bekymringer rundt misbruk, mistolkning, mangel på tid, mangel på ekspertise, mangel på kreditering og tap av kontroll.

Abstract

This master's thesis aims to investigate the attitudes and experiences of researchers in sharing research data. I was interested in learning more about how such data sharing works in practice, and whether there are any specific factors that affect researchers' data sharing. Sharing of research data is encouraged through the EU framework program for research and for research funded through the Research Council of Norway.

I wanted to investigate how researchers respond to these requests, and the challenges they may face when and if they make research data available to others. By interviewing researchers about specific challenges, they have faced and experiences they have gained, I hope to contribute to more insight into the complexity of sharing research data.

I have interviewed twelve researchers, where eight of the researchers have a scientific background and four researchers have a background from the social sciences/humanities. The researchers were mainly positive about sharing data outside their own research group. Although the social science researchers pointed out that there are some limitations in sharing the data from their research. Several of the informants shared the view that data financed with public funds is a public good. For researchers to be willing to share data, there are also certain conditions that must be in place, in addition, the researchers had some concerns related to misuse, misinterpretation, lack of time, lack of expertise, lack of credit and loss of control.



Forord

Arbeidsprosessen knyttet til denne masteroppgaven har til tider vært lang og krevende, men heldigvis også stimulerende. Jeg vil takke min tidligere leder Silje A. Hole, som bifalt og støttet mitt masterprosjekt fra første stund, og min arbeidsgiver IFE (Institutt for energiteknikk) for fleksibiliteten jeg har fått, slik at det var mulig å fullføre mastergraden. En ekstra takk til gode kollegaer for interesse og oppmuntring på veien.

Uten åpenhet og villighet fra informantene mine, ville ikke denne studien ha vært mulig å gjennomføre. Tusen takk til hver enkelt informant som har delt sine tanker og erfaringer med meg.

Jeg vil også takke min veileder Nils Pharo for mange gode råd og innspill. Takk for forståelsen for mitt ønske om å bytte til annen forskningsmetode underveis i prosessen.

Jeg må også rette en takk til min familie, og særlig til min mor – for korrekturlesning nok en gang, og en pep-talk når man trenger det som aller mest.

Til slutt, men ikke minst. Takk for tålmodigheten - Eigil, og for støtte og oppmuntring gjennom hele prosessen.

Innholdsfortegnelse

1. Innledning.....	7
1.1. Bakgrunn og motivasjon.....	9
1.2. Problemstilling.....	9
1.3. Avgrensning.....	10
1.4. Case polarforskning.....	10
1.5. Oppgaven i relasjon til bibliotek- og informasjonsvitenskapelig forskning.....	11
1.6. Oppgavens struktur.....	12
2. Teoretisk perspektiv og konseptuelt rammeverk.....	13
2.1. Kunnskapsinfrastruktur (Knowledge Infrastructure).....	13
2.2. Sosioteknisk perspektiv.....	14
2.3. Hva er forskningsdata?.....	15
2.4. Hva er datadeling?.....	17
2.5. Hva er åpne data?.....	18
2.6. Metadata.....	20
2.7. FAIR-prinsippene.....	20
2.8. Hva er datagjenbruk?.....	22
2.9. Hva kreves for å kunne gjenbruke andres forskningsdata?.....	24
2.9.1. Data, kunnskap og ekspertise.....	24
2.9.2. Forskningsdata som «boundary objects».....	25
3. Bakgrunn og tidligere forskning.....	28
3.1. Veien til åpen forskning (open science).....	28
3.2. EU og åpen forskning.....	29
3.3. Norge – nasjonale prinsipper og retningslinjer.....	29
3.4. Tidligere forskning og litteraturgjennomgang.....	30
4. Metode.....	36
4.1. Vitenskapsteoretisk perspektiv.....	36
4.2. Metodologiske overveielser.....	38
4.3. Kvalitative intervjuer.....	39
4.4. Utvalg av informanter og gjennomføring av intervjuer.....	40
4.5. Strategisk utvalg.....	40

4.6. Rekruttering av informanter	42
4.7. Intervjuguiden.....	42
4.8. Pilot.....	43
4.9 Gjennomføring av intervjuene	43
4.10. Forskningsetikk.....	45
4.11. Konfidensialitet	45
4.12. Bearbeiding av datamaterialet.....	46
4.13. Forskerens rolle.....	46
4.14. Refleksjon.....	47
5. Resultater og analyse	48
5.1. Kort presentasjon av forskernes ulike fagdisipliner/fagfelt.....	50
5.1.1. Naturvitenskap.....	50
5.1.2. Geofysikk og geovitenskap.....	50
5.1.3. Biovitenskap - marinbiologi og toksikologi	55
5.1.4. Samfunnsvitenskap - statsvitenskap og sosialantropologi	60
5.1.5. Humaniora – arkeologi.....	63
5.2. Egenskaper ved dataene	67
5.2.1. Big data vs. «den lange halen av forskningsdata»	67
5.2.2. Kvalitative kontra kvantitative data	70
5.3. Holdninger og erfaringer til deling av forskningsdata utenfor egen forskningsgruppe.	73
5.3.1. Førsteretten til publisering	74
5.3.2. Forskningsdata – forskerens kapital	76
5.3.3. Gjensidighet	78
5.3.4. Tid til å kvalitetssikre dataene	80
5.3.5. Krav om datadeling fra tidsskrifter	81
5.3.6. Hvilke data deles ikke?.....	82
5.3.7. Dataenes «mobilitet».....	84
5.3.8. Dataskaperens fordel	87
5.3.9. Forskningsdata – hva slags produkt skal man dele?.....	88
5.3.10. Datadeling på forespørsel	89
5.3.11. FAIR-data.....	90
5.3.12. Mangel på tid og ressurser	93
5.3.13. Viktigheten av anerkjennelse og kreditering, og mangel på insentiver	94

5.3.14. Hvorfor deles data?.....	96
6. Oppsummering av funn og diskusjon	100
6.1. Sammenligning og kobling mot problemstilling og forskningsspørsmål	100
6.1.1. Datadelingspraksiser	101
6.1.2. Geofysikk/geovitenskap.....	101
6.1.3. Biovitenskap.....	102
6.1.4. Samfunnsvitenskap/humaniora.....	103
6.1.5. Avgjørende faktorer for datadeling	104
6.1.6. Holdninger.....	107
6.2. Avsluttende diskusjon	108
6.2.1. Sosioteknisk perspektiv.....	110
6.3. Refleksjon og begrensing ved studien	113
6.4. Forslag til videre forskning	113
7. Litteraturliste	116
Vedlegg 1.	125
Vedlegg 2.	128
Vedlegg 3.	132
Vedlegg 4.	137

1. Innledning

Data blir sett på som en verdifull ressurs i dagens kunnskapssamfunn. På grunn av den teknologiske utviklingen har det blitt enklere å dele data. En utvikling som også har bidratt til at data ikke lenger kun ses som et biprodukt av det enkelte forskningsprosjekt, men som noe verdifullt i seg selv og noe som kan inspirere til nye hypoteser og muliggjøre nye innsikter og oppdagelser (Tenopir et al. 2011).

For at slike fremtidige oppdagelser skal være mulig, må ikke bare selve forskningsdataene, i form av publikasjoner, men også de data som ligger til grunn for resultatene gjøres gjenbrukbare, og mulige å gjenfinne. Internettet har åpnet for nye muligheter for spredning og tilgang til nytt vitenskapelig materiale. Dette har blant annet gitt fremvekst til bevegelsen *Open Access*, der formålet har vært å gi åpen tilgang til vitenskapelige publikasjoner, det har igjen blitt etterfulgt av diskusjoner om *Open Data* eller *Open Science*. Disse bevegelsene bygger på vitenskapens ideal om åpen deling. Det sies at deling av forskningsdata gir mange fordeler, blant annet gir det muligheter for videre forskning, gjenskaping av analyser, validering og forskningssamarbeid om komplekse problemstillinger, overflødig datainnsamling kan unngås og forskningsdata kan anvendes på nytt (Tenopir et al., 2011).

Ifølge Borgman er det særlig fire begrunnelser eller rasjonaler som vektlegges som argumentasjon for å dele forskningsdata: *(1) å reprodusere eller verifisere forskning, (2) gjøre offentlig finansiert forskning tilgjengelig for allmennheten, (3) gi andre muligheter til å stille nye spørsmål fra eksisterende data og, (4) fremme ny forskning og innovasjon* (Borgman, 2012, s. 1067).

Norske myndigheter vil, sammen med internasjonale organisasjoner som OECD og EU fremme mer deling og arkivering av offentlige finansierte forskningsdata. Det er særlig to grunner til dette. For det første, kan offentlig finansierte forskningsdata anses som et offentlig gode som bør utnyttes i størst mulig grad og ikke reserveres for den enkelte forsker eller institusjon. Dernest blir det hevdet at bedre utnyttelse av forskningsdata kan styrke kvaliteten og ressursutnyttelse i norsk forskning. Det dreier seg om forskningskvalitet og demokratisering av

kunnskap når forskning og kunnskap blir tilgjengelig på tvers av fagmiljøer, sektorer og land (Forskningsrådet, 2019).

Gjennom EØS-avtalen deltar Norge i EUs rammeprogram for forskning og innovasjon. Programmet er EUs største og viktigste verktøy for å finansiere forskning på europeisk nivå. Deling av forskningsdata i åpne arkiver oppfordres for forskning finansiert blant annet av Horisont Europa og forskning finansiert gjennom Norges forskningsråd. Prinsippet for EU og Norge er at «forskningsdata skal være så åpne som mulig, så lukkede som nødvendig» (European Commission, 2021, Kunnskapsdepartementet, 2017, s. 18).

En rekke studier viser at datadeling og arkivering av forskningsdata kan være forbundet med tekniske, økonomiske, kulturelle og juridiske hindringer. Forskere som ikke deler data har ofte spesifikke grunner for å ikke gjøre det (Tenopir et al., 2011).

Formuleringen «så åpent som mulig, så lukket som nødvendig» viser nettopp til disse avveiningene i alle forskningsprosesser. Målene og tiltakene om åpen forskning er ambisiøse og krevende for forskningsmiljøene, utfordringene, og spørsmålet om hva som er mulig og hva som er nødvendig vil måtte inngå i avveiningene i alle forskningsprosesser. I dette spenningsfeltet av motstridende interesser og hensyntagen må forskere manøvrere, og det er mange ulike dilemmaer som kan oppstå og avveininger som man må ta hensyn til. Det er derfor et behov for å se nærmere på forskeres ulike perspektiver.

Tilgjengeliggjøring av forskningsdata ble lansert som policy av Forskningsrådet i 2014. Før policyen ble lansert, ble det gjort en undersøkelse blant forskere i Norge omkring deres holdninger til tilgjengeliggjøring av forskningsdata. Undersøkelsen viste at forskere flest er positive til å dele data, men i praksis er det ikke mange som åpent tilgjengeliggjør forskningsdata for andre (DAMVAD, 2014).

Selv om mange forskere verdsetter og anerkjenner fordelene ved å dele forskningsdata, er de på et individuelt nivå ofte mer tilbakeholdne med å dele sine data. Selv om den allmenne nytten av datadeling for samfunnet synes å være stor, så kan det for den enkelte forsker synes

mer usikkert hva man tjener på å dele sine data. Det finnes flere ulike grunner til å studere hva som driver forskere til å dele sine data. Et stort antall studier har blitt utført i forhold til hva som hindrer forskere å dele data. Mens det er færre studier som har blitt gjennomført for å undersøke hvilke drivkrefter som får forskere til å dele og gjenbruke data (Van den Eynden & Bishop 2014, Borgman 2012). Christine Borgman (2012) vektlegger at det behøves mye mer forskning innenfor disse områdene. Undersøkelser som har blitt utført om hva som driver forskere til å dele data, er i tillegg ofte kvantitative spørreundersøkelser, noe som gjør det vanskeligere å stille oppfølgingsspørsmål, og få en mer grunnleggende forståelse ved å ta del i mer utdypende resonnement. Det er dessuten gjennomført få undersøkelser i Norge, og norske forhold er heller ikke undersøkt i noen særlig grad.

1.1. Bakgrunn og motivasjon

Dette masterprosjektet har til hensikt å undersøke hvilke tanker og erfaringer forskere har rundt deling av forskningsdata. Jeg er interessert i å få vite mer om hvordan slik datadeling fungerer i praksis, og om det er noen spesifikke faktorer som har innvirkning på forskeres datadeling. Det mangler et klart bilde over hvordan forskere responderer på disse oppfordringene, og de utfordringene og problemene de møter når og hvis de gjør forskningsdata tilgjengelige for andre. Hensikten med denne oppgaven er å se nærmere på forskeres ulike utgangspunkt, motivasjoner og opplevde hindringer for å dele data. Ved å intervju forskere om konkrete utfordringer de har møtt og erfaringer de har gjort seg, håper jeg å bidra til mer innsikt i kompleksiteten rundt deling av forskningsdata.

1.2. Problemstilling

På bakgrunn av dette har jeg formulert følgende problemstilling:

Hvilke holdninger og erfaringer har forskere til deling av forskningsdata?

I tillegg har jeg formulert et forskningsspørsmål som vil være med å utdype problemstillingen.

Forskningsspørsmål:

Hvilke faktorer kan ha innvirkning på forskeres datadelingspraksiser?

Med forskningsspørsmålet ønsker jeg å se nærmere på bakenforliggende årsaker som ulike forhold eller faktorer som kommer frem i intervjuene med informantene i denne studien, og som kan tenkes å ha innvirkning på forskeres praksiser med og holdninger til deling av forskningsdata. I begrepet *datadelingspraksiser* – legger jeg forskernes praksiser med å dele egne forskningsdata med andre. I begrepet *praksis* refererer jeg til den aktuelle handlingen med å dele data, heller enn interessen eller intensjonen med å dele data (Knorr-Cetina, 1999).

Holdninger kan defineres som en positiv eller negativ vurdering rettet mot noe eller noen (Ajzen, 2005). Ifølge Kim & Zhang vil holdninger til deling av forskningsdata kunne påvirkes av i hvilken grad en forsker gjør en fordelaktig eller ufordelaktig vurdering av en datadelingsatferd. I tillegg kan ulike faktorer eller forhold ha innvirkning på forskeres datadelingspraksiser. Det kan for eksempel dreie seg om tilgjengelig infrastruktur, tilgjengelighet til dataarkiv/databaser og tilgang på ekspertise. Andre faktorer kan for eksempel dreie seg om føringer eller policyer for datadeling, samt press fra ulike aktører om å tilgjengeliggjøre data (Kim & Zhang, 2015, s. 191) for å nevne noe.

1.3. Avgrensning

En avgrensning som må gjøres er at det kan tenkes andre faktorer som har innvirkning på forskeres datadelingspraksiser og som ikke kommer frem i intervjuene med informantene. Mine funn vil derfor ikke presentere noen uttømmende liste over faktorer som kan tenkes å virke inn på forskernes datadelingspraksiser. Det kan også tenkes at intervjuer med forskere fra andre fagdisipliner, enn de som er representert her ville ha avdekket andre funn.

1.4. Case polarforskning

Valget av informanter til denne studien ble gjort ved å velge ut forskere som har deltatt i forskningsprosjekter innenfor polarforskning. Caset er valgt ut ifra et strategisk utgangspunkt om å få tak i forskere som kunne tenkes å ha erfaring med datadeling, siden forskning utført i polare strøk kan være spesielt ressurs- og kostnadskrevende, og forskere gjerne deler på kostbart utstyr. Fra Forskningsrådets side er det nevnt at deling av forskningsdata er spesielt

viktig innenfor dette forskningsfeltet (Forskningsrådet, 2020, s. 19). Så derfor kunne det tenkes at en del forskere med tilknytning til polarforskning har førstehåndserfaring med å dele forskningsdata. Jeg syntes det ville være interessant å høre om forskeres faktiske erfaringer med å dele data.

Polarforskning er et forskningsfelt som består av forskere med bakgrunn i ulike fagdisipliner og fagområder. En avgrensning som må gjøres er at denne studien ikke vil kunne si noe om hva som kjennetegner forskeres datadeling med tilknytning til polarforskning spesielt. En slik studie måtte ha sett annerledes ut.

1.5. Oppgaven i relasjon til bibliotek- og informasjonsvitenskapelig forskning

I bibliotek- og informasjonsvitenskap har vitenskapelig kommunikasjon blitt studert ut ifra ulike utgangspunkt og med søkelys på ulike deler av forskeres kommunikasjon og bruk av informasjon. Forskningsdata som en del av vitenskapelig kommunikasjon kan innenfor rammen av informasjonsvitenskap som forskningsfelt undersøkes ut ifra en rekke ulike innfallsvinkler og perspektiv, som for eksempel forskningsbibliotekets rolle og deres støttetjenester i forbindelse med forskningsdata, politiske føringer fra myndigheter, teknologiske systemer, eller som i mitt tilfelle ut ifra et forsker- og brukerperspektiv. I tillegg kan man se nærmere på ulike aktører innenfor feltet eller fokusere på aktiviteter innen ulike faser av lagring, forvaltning og deling av forskningsdata.

Et syn på forskningsdata er at det ikke kan betraktes som hver enkelt forskers eiendom, men noe som bør deles med forskersamfunnet og samfunnet for øvrig, dette er også forenlig med Robert K. Mertons (1973) fire normer, som senere ble kjent under akronymet CUDOS¹. Spesielt gjelder det normen om «kommunisme» - som handler om at vitenskapelig kunnskap er et felles gode, og at vitenskapelige fremskritt er betinget av at man offentliggjør og deler resultatene. Forskningsresultater skal dessuten testes og granskes objektivt av andre enn forskeren selv, innen de aksepteres, slik at feil og misforståelser kan avdekkes. Dette tankegodset kan man

¹ CUDOS er en forkortelse for normene *Communism, Universalism, Disinterestedness* og *Organized Scepticism* (Merton, 1973).

finne igjen i dagens forskningspolitikk, og i retningslinjer som omhandler deling av forskningsdata.

Merton hevdet at forskersamfunnet ble styrt av verdier og normer, som oppfattes som bindende av forskerne og forskersamfunnet. Mertons normer har blitt kritisert for å være for funksjonalistiske og for å forenkle vitenskapens mekanismer til et selvregulerende system av normer, verdier og belønninger uavhengig av sosiale faktorer (Van House, 2004, s. 7). Mange senere studier innen vitenskapelig kommunikasjon utgår fra et mer konstruktivistisk perspektiv, og har vist til at forskeres praksiser – heller enn å styres av normer – er sosiale, lokalt situerte og svært varierende (Borgman 2007).

Etter hvert som teknologien gjorde det enklere å skape og dele forskningsdata, viser studier at det å dele data vel så ofte er et sosialt, som et teknisk problem (Van House, 2004). Når data som tidligere var privat, blir offentlig tilgjengelig, reiser det seg en rekke spørsmål og dilemmaer.

1.6. Oppgavens struktur

Masteroppgaven er inndelt i seks kapitler. Innledningsvis har jeg presentert bakteppe for masteroppgaven, motivasjon og problemstilling. I kapittel 2 vil jeg presentere teoretisk perspektiv og konseptuelt rammeverk, samt begreper slik jeg vil anvende dem i denne oppgaven. I kapittel 3 vil jeg presentere bakgrunn og tidligere forskning. I kapittel 4 vil jeg gjøre rede for metoden som er brukt i denne undersøkelsen. I kapittel 5 blir resultater fra undersøkelsen presentert og analysert. I kapittel 6 avrundes oppgaven med en oppsummering av funn, og resultater kobles opp mot problemstilling og forskningsspørsmål, i tillegg vil jeg se nærmere på deling av forskningsdata ut ifra et sosioteknisk perspektiv. Til slutt gis det anbefalinger for videre forskning.

2. Teoretisk perspektiv og konseptuelt rammeverk

Følgende kapittel gir en introduksjon til og overblikk over noen konsepter som er viktige å kjenne til innenfor en studie om deling av forskningsdata. Jeg vil også introdusere noen definisjoner og nøkkelbegreper, og redegjøre for hvordan jeg kommer til å anvende dem i denne oppgaven.

2.1. Kunnskapsinfrastruktur (Knowledge Infrastructure)

Borgman hevder at data ikke har noen verdi eller mening i seg selv isolert sett, men at data eksisterer innenfor en kunnskapsinfrastruktur – der komplekse nettverk av mennesker, praksiser, teknologier, institusjoner, materielle objekter og relasjoner inngår og er sammenvevd med hverandre. Det kreves mange slags ekspertkunnskaper for å håndtere data i ulike stadier av forskningsprosessen: teoretiske, praktiske, sosiale og tekniske (Borgman 2015, s. 36).

Infrastrukturer er nødvendige for å kunne samle inn, lagre og bruke data, siden disse aktivitetene er en del av vitenskapelig praksis (Bowker, 2005). I denne oppgaven refererer begrepet «knowledge infrastructure» (KI), eller kunnskapsinfrastruktur til det komplekse sosio-tekniske miljøet hvor vitenskapelig kunnskap blir produsert, delt, brukt og gjenbrukt (Borgman et al., 2016). Jeg anvender Edwards definisjon av KI: «robust networks of people, artifacts, and institutions that generate, share and maintain specific knowledge about human and natural worlds» (Edwards, 2010, s.17).

Under denne definisjonen inkluderer kunnskapsinfrastrukturer foruten teknologiske løsninger, også individer, organisasjoner, rutiner, delte normer og praksiser. Infrastrukturer er ikke konstruerte eller fullstendig sammenhengende prosesser. Snarere forstås de best som økosystemer eller komplekse adaptive systemer. De består av mange deler som samhandler gjennom sosiale og tekniske prosesser, hver med unike opprinnelser og mål, som er laget for å fungere sammen ved hjelp av standarder, protokoller, sosiale praksiser, normer og individuell adferd som «smører» forbindelsene mellom dem. Denne tilpasningsprosessen er kontinuerlig, ettersom individuelle elementer endres og nye introduseres – og den er ikke nødvendigvis alltid vellykket (Borgman et al., 2019, Borgman 2015, Edwards et al., 2013).

Muligheten for å generere eller samle inn, prosessere og utveksle datasett avhenger ikke bare av vitenskapelig ekspertise, men er også avhengig av infrastruktur som gjør det mulig å oppdage, gjenfinne, tolke og bruke datasettene (Borgman, 2015, Edwards et al., 2013).

2.2. Sosioteknisk perspektiv

Når en studerer kunnskapsinfrastrukturer innenfor forskning, tar man i informasjonsvitenskapelige studier ofte utgangspunkt i en «sosioteknisk tilnærming» for å analysere forskeres arbeidspraksiser og deres artefakter, slik som for eksempel notatbøker, datasett, programvare også videre. Begrepet sosioteknisk blir brukt for å tilbakevise idéen om at sosiale og tekniske fenomener er distinkte og motstridene (Star, 1995). Det sosiotekniske perspektivet forsøker å beskrive samspillet mellom de sosiale og de tekniske aspektene i et system, samt deres interaksjon med omverden. De ulike delene bygger til sammen opp systemet, og påvirker hverandre gjensidig. Ingen deler av systemet, sosiale eller tekniske, kan ses som isolerte eller upåvirkede av andre deler (Golden, 2013). Ifølge Golden er teorien tydelig på at begrepet «teknisk» ikke begrenses til å beskrive ulike enheter eller maskiner, men heller brukes i en videre betydning til å beskrive arbeidsprosesser, så vel som ulike typer utstyr. Likeledes, brukes begrepet «sosial» ikke bare for å beskrive hvordan enheter, maskiner eller arbeidsflyt påvirker et individ, men også hvordan det kan påvirke ferdigheter, kunnskap, holdninger, sosiale relasjoner, og nettverk av forbindelser og interaksjoner mellom individer og grupper. (Golden, 2013).

Et sosioteknisk system skapes og opprettholdes i samspill mellom mennesker, teknikk og praksiser, der teknologien setter rammene for hva mennesker kan gjøre, samtidig som menneskers behov og adferd, driver frem en stadig utvikling av teknologien.

Sosioteknisk perspektiv brukes ofte for å beskrive det komplekse samspillet mellom mennesker og teknologi, der hverken det sosiale (mennesker, relasjoner, strukturer etc.) eller det teknologiske (enheter, utstyr, prosesser, materialer etc.) kan bli betraktet isolert fra hverandre dersom ytelsen skal bli optimalisert (Golden, 2013). Det er sjelden at disse faktorene kan skilles fra hverandre, siden de påvirker hverandre gjensidig (Borgman, 2015).

Bowker et al. poengterer, at når man studerer vitenskapelige praksiser, istedenfor å vurdere om et problem først og fremst er av teknisk eller sosial art, bør man heller fokusere på om løsningen for et gitt problem er enten primært sosialt, teknisk eller en kombinasjon av begge deler (Bowker et al. 2009, s. 102).

2.3. Hva er forskningsdata?

Ifølge Johannessen et al. betyr ordet data – flertall av det latinske *datum* – noe som er gitt. Når virkeligheten iakttas og undersøkes, og det på en eller annen måte registreres, er virkeligheten blitt data. Når det er sagt, så er data mye mer enn det. Det er også mulig å registrere fenomener som ikke umiddelbart fremstår som iøynefallende, eller som kan betraktes med det blotte øye. Data er mer eller mindre vellykkede *representasjoner* av virkeligheten. Men selv den mest nøyaktige iakttagelse og gjengivelse kan aldri fange inn den autentiske virkeligheten, men kun et utsnitt eller en bestanddel av det som studeres (Johannessen et al., 2016, s. 32).

Forskningsdata kan være i enten fysiske eller digitale formater. Fysiske forskningsdata kan utgjøre det meste mellom himmel og jord. Noen data blir digitale først etter at forskere bearbeider fysiske data, mens andre data er «født» digitale. Digitale data kan se ut på mange ulike måter, de kan for eksempel være tekstbaserte, numeriske, relative eller instrumentgenererte (Kennan 2017). Data kan også være skapt ved hjelp av ulike metoder – man snakker for eksempel gjerne om observasjonsdata, eksperimentelle data og modellbaserte data (Borgman, 2010). Data kan for eksempel også være i form av bilder, film eller lyd. De kan være innsamlet eller skapt av mennesker, så vel som maskiner - og kreve ulike instrument eller programvare for å være anvendbare (Borgman, 2015).

Det finnes ingen entydig definisjon av begrepet forskningsdata, og begrepet forstås gjerne ulikt innenfor forskjellige fagdisipliner og forskningsmiljøer (Forskningsrådet, 2021). Avhengig av konteksten kan data være mange ting. Hvilke aspekter man legger i begrepet vil variere ut ifra hvilket utgangspunkt man har, og hvilken fagdisiplin eller forskningsmiljø man tilhører. Ulike fagdisipliner har ulike behov og syn på hva data er, og hvordan data skal håndteres (Borgman, 2015, s. 55-56).

Borgman definerer data som «representations of observations, objects or other entities used as evidence of phenomena for the purpose of research or scholarship” (Borgman, 2015, s. 29).

Denne definisjonen er anvendelig for å kunne forklare når en observasjon eller andre former for registreringer eller representasjoner blir til data. Det hjelper også til med å forklare hvorfor data blir til data, først i perspektivet til den som betrakter informasjonen som data. Borgman mener at «data are not pure or natural objects with an essence of their own. They exist in a context, taking on meaning from that context and from the perspective of the beholder” (Borgman, 2015 s. 18). Data er gjenstand for tolkning, statusen som fakta eller bevismateriale bestemmes av personene som produserer, administrerer og bruker disse dataene (Borgman, 2007, s. 119-121).

Forskningsrådet (2021) henviser til blant annet OpenAIRE² sin definisjon av forskningsdata:

“Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others. Data may be defined as ‘relational’ or ‘functional’ components of research, thus signalling that their identification and value lies in whether and how researchers use them as evidence for claims. They may include, for example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript” (OpenAIRE, 2021).

Forskningsrådet påpeker at det finnes mange definisjoner av forskningsdata, men en fellesnevner er at forskningsdata er informasjon som på en eller annen måte brukes eller oppstår i forbindelse med forskningsaktivitet. De fleste definisjoner trekker frem at data er informasjon som brukes som underlag eller grunnlag for vitenskapelige funn eller analyser, og

² <https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected>

at det også kan brukes for å etterprøve og validere resultater fra forskning (Forskningsrådet, 2021, s. 11). Selv om det siste ikke kan sies å være aktuelt eller tilfelle i alle forskningsdisipliner. Som Borgman, oppsummerer det eneste man er enig om er at ingen definisjon er tilstrekkelig for å romme hva forskningsdata er i enhver sammenheng (Borgman, 2015, s. 4).

2.4. Hva er datadeling?

Datadeling refererer vanligvis til handlingen å gjøre data tilgjengelig, slik at de kan gjenbrukes av andre. Borgman har definert deling av forskningsdata som «the release of research data for use by others» (Borgman, 2012, s. 1060).

Når forskere snakker om datadeling, mener de ofte en rekke forskjellige måter forskningsdata blir utvekslet med andre forskere på (Van den Eynden & Bishop, 2014). Deling av data kan ta mange ulike former, og det kan skje med ulike grader av formalitet og tilgjengelighet, fra privat utveksling av data på forespørsel forskere imellom, til deponering av datasett i dataarkiv med tilgang for offentligheten (Borgman, 2012, Whyte & Pryor, 2011).

Datadeling omfatter derfor mange måter å tilgjengeliggjøre data på, og sier lite om hvor gjenbrukbare disse dataene er. Eksempler på datadeling inkluderer privat utveksling av data forskere imellom, å poste et datasett på en forskers eller et forskningsprosjekts nettside, deponere datasett til en database/dataarkiv eller en fagspesifikk samling, eller legge ved data som vedlegg til en tidsskriftsartikkel (Wallis et al. 2013). En relativt nyere praksis i noen fagfelt er å dele datasett som en dataartikkel³. Dataartikler gir beskrivelser om hvordan datasett er innsamlet, hvordan dataene er behandlet, og for eksempel hva slags programvare som er brukt. Dette øker muligheten for dataproveniens og det er en måte for dataprodusentene/forfatterne å få anerkjennelse eller bli kreditert for data som de selv har generert eller samlet inn (Pasquetto et al., 2017, s. 2).

Whyte & Pryor (2011) har sett nærmere på to dimensjoner i forbindelse med deling av forskningsdata. Den ene dimensjonen er fasen i forskningsprosessen der data deles, fra råmateriale i den ene enden, til publiserte artikler og datasett i den andre. Den andre

³ På engelsk *data paper*. Beskriver en eller flere datasett og mulige måter å anvende dem på, men inneholder normalt ikke forskningsresultater, analyser eller konklusjoner.

dimensjonen går på involverte aktører og grad av åpenhet i den andre enden. Ut ifra dette definerer Whyte & Pryor seks ulike nivå av åpenhet og måter å dele data på:

1) *Private management* — data deles innenfor forskergruppen, 2) *Collaborative sharing* – data deles innenfor et konsortium (forskningsprosjekt/program), 3) *Peer exchange* – data deles med fagfeller i uformelle nettverk, 4) *Transparent governance* – eksterne parter får tilgang for å kunne granske og vurdere forskningen, 5) *Community sharing* – tilgang begrenses til medlemmer i forskersamfunnet, 6) *Public sharing* – allmennheten får tilgang til dataene.

Konseptet med å dele data kan forstås på ulike måter av ulike aktører. Det er ikke etablert en felles akseptert forståelse for begrepet «deling av forskningsdata», selv om det blant forskningsråd, forlag, tidsskrifter og myndigheter oppmuntres til deling av forskningsdata, der offentlige dataarkiv eller databaser blir sett på som det foretrukne middelet for å tilgjengeliggjøre forskningsdata (Borgman, 2019, Thoegersen & Borlund, 2022).

Metoder for datadeling kan variere med fagdisipliner, datatyper, land, tidsskrifter, hvordan forskningen blir finansiert og andre faktorer. Muligheten for å oppdage, gjenfinne og tolke delte data vil også variere tilsvarende (Borgman, 2015, Pasquetto et al., 2017).

2.5. Hva er åpne data?

Som vi har sett kan begrepet «deling av forskningsdata» referere til en rekke ulike måter å dele data på. I noen sammenhenger brukes det synonymt med «open data» eller «åpne data» (Thoegersen & Borlund, 2022). Det er liten grad av konsensus når det kommer til å enes om hva som gjør data åpne (Borgman, 2015, s. 44).

Det er imidlertid nødvendig å differensiere mellom «deling av forskningsdata» og begrepet «åpne data». «Åpne data» er et problematisk begrep, gitt spekteret av konsepter og forhold det kan referere til. Til tross for mangelen på enighet rundt begrepet «åpne data», er det vanlig å referere til noen grunnleggende betingelser for å gjøre forskningsdata «åpne». Det refererer vanligvis til å gjøre data tilgjengelig med «færrest mulige restriksjoner» og «til lavest mulige kostnader» (Pasquetto, et al. 2017).

I OECD- rapporten: Principles and Guidelines for access to research data from public funding (2007), legges det til grunn at forskningsdata finansiert med offentlige midler skal være tilgjengelig for et internasjonalt forskerkollektiv, - på like betingelser, og til lavest mulig kostnad. Tilgangen skal dessuten være enkel, innen rimelig tid (timely), brukervennlig, og hvis mulig internettbasert (Organisation for Economic Co-operation and Development, 2007, s. 15).

Det er ikke heller alltid nødvendigvis enighet om de tiltenkte målgruppene for åpne data. Mens noen politiske organisasjoner (som OECD) fokuserer på idéen om at data hovedsakelig bør være åpne for forskersamfunnet, inkluderer andre forkjempere (som Open Knowledge Foundation⁴) for åpen vitenskap – allmennheten.

Borgman påpeker at åpenhet til data kan variere på mange måter. Offentlige databaser kan tillate bidragsytere å beholde copyright og kontroll over dataene de har deponert. Data kan være åpne, men kun tolkbare med proprietær programvare. Data kan være generert med åpen kildeprogramvare, men samtidig kreve lisens for datagjenbruk (Borgman, 2015). Å ivareta eller kuratere data over en lang tidsperiode krever ofte kontinuerlige investeringer i kuratering for å tilpasse seg endringer, som for eksempel til brukergruppen eller nye tekniske oppdateringer (Baker et al. 2015).

I Forskningsrådets policy *Tilgjengeliggjøring av forskningsdata* fra 2017, så var hovedprinsippet «åpen som standard» retningsgivende for tilgjengeliggjøring av forskningsdata. Forskningsrådet (2021) skriver at i en tidligere fase var det om å gjøre å få med seg alle miljøer på idéen om at deling av forskningsdata var essensielt og riktig i moderne forskning, og man innførte derfor konseptet «åpen som standard». Dette har senere ført med seg en del kritikk, siden det i liten grad tar hensyn til utfordringer knyttet til personvern, sikkerhet, kommersialisering og kostnader relatert til å gjøre data åpne. Konseptet har med tiden blitt noe moderert, og derfor snakker man nå heller om å gjøre data «så åpne som mulig og så lukket som nødvendig» (Forskningsrådet, 2021, s. 13).

⁴ "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)." Kilde: <http://opendefinition.org/>

2.6. Metadata

Ifølge Forskningsrådet (2017) skal forskningsdata utstyres med standardiserte metadata, dette vil gjøre andre i stand til å søke etter og ta i bruk dataene. De bør gi en beskrivelse av datakvaliteten og følge internasjonale regler. Metadataene bør dessuten gjøres tilgjengelig uten kostnad og publiseres slik at de kan høstes maskinelt og brukes i søk etter forskningsdata (Forskningsrådet, 2017, s. 6).

Som vi har sett understrekes betydningen av gode metadata, dersom forskningsdata skal kunne gjenbrukes. Kort forklart er metadata, "data om data", det vil si informasjon som beskriver data og gir mening til data. Metadata er strukturert informasjon som beskriver, forklarer, lokaliserer eller på annet vis gjør det enklere å gjenfinne, bruke eller forvalte en informasjonsressurs (NISO 2004).

2.7. FAIR-prinsippene

Det har vært økende oppmerksomhet rundt manglende infrastruktur for deling av forskningsdata, og et økende fokus for å forbedre infrastrukturen som støtter opp om gjenbruk av forskningsdata. Dette førte til at et mangfoldig utvalg av ulike interessenter (fra blant annet akademia, industri, finansieringsinstitusjoner og vitenskapelige forlag) kom sammen for å designe og gi felles støtte og oppslutning til et sett av prinsipper som kalles «The FAIR data principles» (Wilkinson et al. 2016).

Hensikten er at disse prinsippene skal fungere som en veiledning for å gjøre forskningsdata mest mulig tilgjengelige. Det har også med tiden blitt stilt større krav fra myndigheter, finansører, institusjoner og utgivere at (offentlige) forskningsprosjekter skal ha en dataforvaltningsplan og at forskerne skal følge FAIR-prinsippene så langt det lar seg gjøre. FAIR-prinsippene er ment å være retningslinjer, som skal være til hjelp for å redusere tekniske og administrative hindre for tilgang og gjenbruk, slik at data kan tilfredsstille ulike kvalitetskrav. FAIR-prinsippene legger også stor vekt på at data befinner seg i et digitalt økosystem, og at de må kunne håndteres maskinelt, mest mulig uavhengig av mennesker (Wilkinson et al. 2016).

FAIR er et akronym som bygger på fire grunnleggende prinsipper, disse er: Findable, Accessible, Interoperable og Reusable (Wilkinson et al. 2016). Disse kan oversettes på norsk til: gjenfinnbar, tilgjengelig, interoperable og gjenbrukbar.

For at data skal være gjenfinnbare, må de for eksempel være utstyrt med en unik og bestandig identifikator (Persistent Identifier, eller PID). Dette vil gjøre det mulig å referere og peke til dem, i praksis vil dette ofte si et DOI-nummer. En DOI (Digital Object Identifier) er en varig og unik identifikator. Dette legger til rette for en korrekt sitering av forskningsdata, samtidig som det bidrar til verifikasjon og gjenbruk. DOI bør finnes tilgjengelig i dataarkiv og i andre søketjenester. Data må også beskrives med rikholdig metadata, som gjør det mulig å oppdage dataene ut ifra deres kjennetegn (Wilkinson et al., 2016).

For at data skal være tilgjengelige må de være lett å få tilgang til. For å oppnå det bør applikasjonen tilby en veldefinert og åpen protokoll for data- og metadatautveksling. Protokollen må tillate en autentisering eller autorisasjonsprosedyre dersom det er nødvendig. Metadata skal være tilgjengelige, selv om dataene ikke lenger er tilgjengelige (Wilkinson et al, 2016).

I begrepet interoperable ligger det at både data og metadata skal kunne håndteres maskinelt, og at man skal bruke konsistente vokabular som følger FAIR-prinsippene. Slik kan også metadata og selve dataene fra ulike prosjekt utveksles med hverandre, slik at det går an å søke, sammenligne, samt behandle data og metadata på tvers av teknologiske systemer. Derfor må metadata være i et formelt, tilgjengelig og velegnet språk (Wilkinson et al. 2016).

For at dataene skal være gjenbrukbare må de være tilstrekkelig dokumentert, og det må registreres så fullstendige metadata som mulig. Metadataene bør imøtekomme domene-relevante fellesstandarder. Det må også følge med en detaljert proveniens. Denne kan for eksempel inneholde informasjon om hvordan dataene er samlet inn, samt hva slags programvare og instrumenter som er anvendt. Hva slags programvare som er nødvendig for å åpne filene, og hvordan dataene er kodet og så videre. Hvis dette ikke følger med, vil dataene som regel være umulige for andre å bruke. Derfor må alt som er relevant for videre bruk være beskrevet. Et annet spørsmål, som relaterer seg til gjenbruk, er hva andre kan gjøre med de

dataene som er gjort tilgjengelige. Det må derfor knyttes klare brukslisenser til dataene. En lisens vil avklare hvordan data kan bli gjenbrukt (Wilkinson et al, 2016).

Ifølge Mons et al. har det vært stor oppslutning rundt FAIR-prinsippene de siste årene. Flere internasjonale organisasjoner og land viser til dem som retningsgivende for hvordan man bør håndtere forskningsdata. Det har imidlertid oppstått ulike tolkninger og oppfatninger knyttet til FAIR-prinsippene. Forskningsdata skal som hovedprinsipp være åpne for å sikre gjenbruk og vitenskapelige fremskritt. Allikevel er det en misforståelse når noen tolker det som at all data skal være åpen. Ifølge Mons et al. står A' en i FAIR for «Accessible under well defined conditions» (Mons et al., 2017, s. 51).

Det kan være legitime årsaker til at data ikke er åpent tilgjengelig. I enkelte tilfeller kan det være nødvendig å skjerme data fra allmennheten, for eksempel der åpne data kan tenkes å bidra til at sikkerheten til mennesker eller arter blir truet. Hensyn til for eksempel personvern og kommersialisering gjør også at man ikke kan offentliggjøre alle data. Et grunnprinsipp er som tidligere nevnt at forskningsdata skal være «så åpne som mulig, og så lukkede som nødvendig». Dette beskriver også FAIR-prinsippene «the A allows fair shielding or protection of data that cannot be open (...). On the other hand, from the basic principle the FAIRness is maximised when data are open, maximising A implies maximising openness» (Mons et al, 2017, s. 52).

I EU-kommisjonens rapport *Progress on Open Science: Towards a Shared Research Knowledge System* (2020), understrekes det at FAIR prinsippene bør være målet, og ikke nødvendigvis at dataene skal være åpne. Rapporten etterlyser også blant annet en økende bevissthet rundt verdien av «immaterielle rettigheter» (intellectual property rights) (European Commission, 2020).

2.8. Hva er datagjenbruk?

Det er ikke etablert en felles forståelse eller noen standarddefinisjon av begrepet «datagjenbruk» (Van de Sandt et al., 2019). Datagjenbruk kan bety mange ulike ting for ulike personer. Paschetto et al., viser til at datagjenbruk kan være et vidt konsept som favner mange ulike aktiviteter, som for eksempel å returnere til ens egne data for senere sammenligninger.

Det kan for eksempel også være å innhente datasett fra offentlige eller private kilder for å sammenligne med nylig innsamlede data, kartlegging av tilgjengelige datasett som bakgrunnsforskning for et nytt prosjekt, eller utføre reanalyser av ett eller flere datasett for å adressere nye forskningsspørsmål (Pasquetto et al., 2019, s. 4). Forskere kan gjenbruke andres data for mange ulike formål.

Den mest fundamentale utfordringen med å definere datagjenbruk er å skille mellom «bruk» og «gjenbruk», skillet vil være av både tidsmessig og kontekstuell art. Gjenbruk av data er en aktivitet som kan oppstå over en lang tidshorison, og derfor er det også vanskelig å måle denne aktiviteten. I situasjoner hvor data er samlet inn av et individ eller en forskningsgruppe, vil den første anvendelsen av data vanligvis være når et individ eller en forskergruppe utforsker det i forhold til et spesifikt forskningsspørsmål. Hvis den samme personen eller forskningsgruppen returnerer til ens egne data ved en senere anledning, vil det være en annen bruk av dataene, og slik sett kan det betegnes som gjenbruk, men dataene er fortsatt en del av opphavspersonens eller forskningsgruppens kontekst (Pasquetto et al., 2017, s. 3).

Dette er imidlertid ikke den type gjenbruk jeg har hatt til hensikt å studere når jeg har spurt mine informanter om datadeling og gjenbruk i min undersøkelse. Når det gjelder datadeling har jeg lagt Borgmans definisjon til grunn for denne oppgaven: «the release of research data for use by others» (Borgman, 2012, s. 1060). Med utgangspunkt i Borgmans definisjon for deling av forskningsdata, så fulgte det naturlig å legge seg på samme linje for gjenbruk av data. Med utgangspunkt i Borgmans definisjon for datadeling, vil jeg definere gjenbruk i min oppgave til: «to (re)use research data released by others».

I denne oppgaven innebærer gjenbruk – at data blir brukt av noen andre, enn den opprinnelige opphavspersonen (eller den opphavelige forskningsgruppen). Pasquetto et al. hevder at i den vanlige dagligtalen (common parlance), når man snakker om forskeres datapraksiser, så impliserer gjenbruk vanligvis «usage of a dataset by someone other than the originator» (Pasquetto et al., 2017, s. 3).

2.9. Hva kreves for å kunne gjenbruke andres forskningsdata?

Borgman påpeker at muligheten for å gjenbruke andres forskningsdata avhenger av kunnskap om disse dataene og kjennskap til hvordan de har blitt skapt. Det er dataskaperne selv som vanligvis har den mest fortrolige eller inngående kunnskapen om et datasett. Denne kunnskapen blir til gjennom planlegging, innsamling, prosessering, analyse og tolkning av dataene. Borgman peker også på at det kan være mange individer som er involvert i genereringen av data, slik at kunnskap om dataene kan være spredt blant flere parter på en gang. Senere gjenbrukere av datasett vil søke den kunnskapen som trengs gjennom metadata, dokumentasjon, kontakt med dataskaperne, eller andre måter som vil være nødvendige for å forstå dataene (Borgman, 2015). Hvilken type kunnskap som trengs for å kunne gjenbruke og forstå data frembrakt av andre, avhenger også av gjenbrukerens avstand med opprinnelsen til dataene, enten det er avstand i tid, teori, fagdisiplin, egen ekspertise, ressurser eller andre faktorer, - i tillegg til det tiltenkte formålet med å gjenbruke dataene (Borgman, 2015).

2.9.1. Data, kunnskap og ekspertise

Det finnes flere måter å kategorisere kunnskap på i litteraturen. Den mest tradisjonelle er å skille mellom data, informasjon og kunnskap (Gottschalk, 2003). Kirchner skiller begrepet kunnskap fra «data» (uprosesserte fakta) og «informasjon» (meningsfull samling av data), ved at kunnskap involverer en persons tolkning, ferdigheter og erfaringer for å bearbeide informasjonen og konvertere den til kunnskap (Kirchner, 1997). Kunnskap er derfor informasjon som har blitt tolket og bearbeidet av den som skal anvende informasjonen. Davenport & Prusak mener i tillegg at kunnskap er en flytende miks av erfaringer, verdier, kontekstuell informasjon, ekspertinnsikt, som gir grunnlag for en forståelsesramme for å evaluere og inkorporere nye erfaringer – og som igjen danner utgangspunkt for ny kunnskap i en persons sinn (Davenport & Prusak, 1998).

Pasquetto et al., mener at forskjeller mellom kunnskap og ekspertise, et av kjerneproblemene i epistemologien, hjelper til med å forklare datagjenbruk (Pasquetto et al., 2019). Den mest essensielle forskjellen er mellom «knowledge that» og «knowledge how» (Ryle, 1949). «Knowledge that» vil si det å ha kunnskap om noe (episteme) og «knowledge how» vil si det å

ha praktisk kunnskap (tekne). Kunnskapen som skal til for å forene disse og vite når og hvordan de kan brukes kalles *fronesis* (Page & Moen, 2020). En person kan ha «knowledge that», men ikke være i stand til å forklare hvordan noe fungerer. Vi kan lære «knowledge that» fra kodifisert informasjon som artikler eller bøker (Borgman, 2007). «Knowledge how» er en mer avansert form for ekspertise, hvor personen har kunnskap om hvordan noe skal gjøres. Etter hvert som folk blir dyktigere til en oppgave, får de taus kunnskap, en form for ekspertise som er vanskelig å artikulere og uttrykke (Polanyi, 1966). Taus kunnskap er erfaringsbasert og vanskelig å formalisere, og blir derfor ofte overført til andre på mer uformelle og subtile måter. Taus kunnskap kan beskrives med det kjente sitatet «we know more than we can tell» (Polanyi, 1966, s. 4).

Kunnskap er sosialt konstruert. Individuer lærer gjerne «knowledge how» fra praksis i et sosialt fellesskap (Borgman, 2007). Van House har tatt utgangspunkt i Jean Lave og Etienne Wengers syn på praksisfellesskap (community of practice), der læring er tett sammenkoblet med sosiale praksiser, og materielle fenomen, og skriver: «People learn and work within groups that share understanding, practices, technology, artifacts and language» (Van House, 2003, s. 283).

Van House (2003) argumenterer for at det er særlig tre karakteristikk som ligger til grunn for kunnskap. Den er situert, spredt, og sosial:

“That it is situated means that knowledge work is performed by specific people under specific conditions for specific purposes. It is distributed because it entails cooperation among people who don’t know each other and those who do, across space and time. [...] Finally, it is social because we work and learn together and decide what and whom to believe and rely on in the community. Much of what we claim to know comes not from our own direct experience but from what other tell us is so, including our knowledge of whom to believe” (Van House, 2003, s. 271).

2.9.2. Forskningsdata som «boundary objects»

Forskningsdata er bearbeidet for spesifikke forskningsformål. Som en konsekvens er forskningsdata – som alle data – lokale og historisk situerte artefakter (Gitelman, 2013, Star & Griesemer, 1989). Data er ikke naturlige objekter med en egen essens eller mening i seg selv, de eksisterer i en kontekst. Borgman viser til at flere studier med utgangspunkt i

samfunnsvitenskap og filosofi, fortsetter å møte på spenninger mellom evnen til å gjenbruke forskingsdata i ulike sammenhenger og kontekster, og informasjonstapet som oppstår når data fjernes fra deres opprinnelige produksjonskontekst (Borgman, 2015). Data er en fundamental del av forskningsprosessen, med de kan være vanskelige å ekstrahere som produkter som skal deles (Pasquetto et al., 2019). Dette er også kjent som «the mobility problem» (Borgman, 2015, s. 219).

Til forskjell fra vitenskapelige publikasjoner, som er dokument skapt for å kunne spres og stå for seg selv, må data settes inn i en kontekst (Borgman, 2015, Mauhtner & Parry, 2013). Hvis data skal være mulig å gjenbruke, må de organiseres, utstyres med metadata og dokumenteres slik at de kan forstås. Å skape denne konteksten tar tid. Det å dokumentere forskningsdata til bruk for sin egen forskningsgruppe, vil være mer utfordrende enn å dokumentere forskningsdata til eget bruk. Å dokumentere forskningsdata til bruk utenfor forskningsgruppen er enda vanskeligere, og det aller vanskeligste vil være å dokumentere forskningsdata for ukjente fremtidige brukere (Borgman, 2007). Metadata, og ontologier⁵ eller andre former for informasjonsbeskrivelser (som for eksempel standarder), kan være måter å formalisere og overføre slik informasjon. I hvilken grad disse kontekstene og meningene kan representeres - har innflytelse på deres overførbarhet (Borgman 2007, s. 168).

Sabina Leonelli, historiker og vitenskapsfilosof, skiller mellom ontologier som «relevance labels» og metadata som «reliability labels» som situerer og rommer mange «små fakta». Sammen knytter disse mekanismene datasett til spesifikke forskningsobjekter (f.eks. den biologiske enheten som studeres) og gir informasjon om kvaliteten på dataene, slik som dataformat, organismer brukt i eksperimenter, instrumenter og metoder som brukes, og for eksempel under hva slags laboratorieforhold dataene ble innhentet (Leonelli, 2010).

Når de er vellykket anvendt, hjelper metadata og ontologier, forskningsdata til å fungere som «mobile» objekter (Latour, 1987). Det vil si meningsfulle objekter som kan bevege seg mellom ulike brukere eller forskningsprosjekter, samtidig som de beholder tilstrekkelig «beviskraft».

⁵ Ontologier - formell representasjon av et sett begreper (med mellomliggende relasjoner) innenfor et kunnskapsområde. Kilde: Det norske akademis ordbok.

Disse beskrivelsene eller merkelappene muliggjør at datasett kan fungere som «boundary objects» – objekter som er plastiske eller formbare nok til å tilpasse seg lokale forhold, men likevel robuste nok til å opprettholde en felles identitet på tvers av ulike lokasjoner (sites) (Star & Griesemer, 1989, s. 408).

Man kan si at å gjenbruke datasett frembrakt av andre vil være å gi dem status som «boundary objects». Van House et al. mener at å gi dataene status som «boundary objects» vil være å vurdere at dataene er tilstrekkelig robuste nok til å bli brukt på tvers av ulike anvendelsesområder (sites), og formbare nok til å tilpasses lokale behov. Men en potensiell gjenbruker vil også vurdere om vedkommende forstår dataene godt nok til å vite hvordan man kan bruke dem (Van House et al., 1998). Gjenbruk av data innebærer også at gjenbrukeren vurderer kvaliteten og påliteligheten av innholdet. Å vurdere påliteligheten til et datasett er ikke alltid lett (Van House, 2003, s. 278). Kan man stole på eller anta at den opprinnelige dataskaperen har fulgt adekvate vitenskapelige praksiser? Forskeren som skal gjenbruke dataene må først oppnå tillit til at dataene er til å stole på og at de er pålitelige. Denne tilliten blir kontinuerlig vurdert og revurdert av den som vurderer å gjenbruke dataene (Van House et al., 1998). Datagjenbruk involverer frakobling eller en dekontekstualisering av dataene fra forskerne som samlet dem inn. Dette åpner opp for spørsmålet om hvordan data innsamlet eller skapt av en forsker, kan oppnå tillit og bli forstått av en annen (Carlson & Anderson, 2007).

Ifølge Bruno Latour, utvikler forskere dyp ekspertise og kunnskap innenfor sitt forskningsfelt, noe som for eksempel også kan innebære kjennskap til spesifikke metoder og vitenskapelig utstyr, som alle er integrert i dataene de samler inn, analyserer og tolker. Som følge av den kunnskapen som er involvert i dannelsen av dem, kan det være vanskelig å løsrive eller skille dataene fra de kontekstene de oppstod i (Latour, 1987).

3. Bakgrunn og tidligere forskning

Følgende kapittel gir en introduksjon til begrepet *open science* eller *åpen forskning*, som er den betegnelsen som brukes av Norges forskningsråd. Jeg vil også kort presentere sentrale retningslinjer, siden disse er med å sette rammene for feltet. Deretter vil jeg presentere noen tidligere studier og litteratur på feltet som vil være aktuell for denne studien.

3.1. Veien til åpen forskning (open science)

Åpen forskning kan sies å være et paraplybegrep som beskriver pågående endringer for hvordan forskere samarbeider og hvordan forskning gjennomføres, kunnskap deles og vitenskapen organiseres. Åpen forskning kan betraktes som en bevegelse der målet er å tilgjengeliggjøre og åpne opp forskningen, i tillegg til prosessene som er forbundet med forskningen. Åpen forskning representerer «en ny tilnærming til den vitenskapelige prosessen tuftet på samarbeid og nye måter å dele kunnskap på, ved å bruke ny teknologi og nye samarbeidsverktøy» (European Commission, 2016, s. 33).

Denne oppgaven fokuserer på deling av forskningsdata som er et av hovedaspektene ved åpen forskning. Et annet hovedaspekt som inngår i open science-begrepet, er åpen tilgang til vitenskapelige publikasjoner. Andre aspekter ved åpen forskning er åpen kildekode (open source), åpen fagfelleevaluering (open peer review), åpent samarbeid (open collaboration) og folkeforskning (citizen science) for å nevne noen eksempler på det som hører inn under begrepet åpen forskning (Forskningsrådet, 2019).

Det understrekes en gang iblant at den grunnleggende idéen om åpenhet innen forskning og vitenskap ikke er noe nytt fenomen. Det kan snarere sies å være et grunnprinsipp for den moderne vitenskapen, der åpen kommunikasjon og granskning av idéer, resultat og data har vært avgjørende for vitenskapelige fremskritt. Åpen forskning kan sies å være et møte mellom denne gamle tradisjonen, og den teknologiske utviklingen. Der fremveksten av internett og nye digitale løsninger har skapt nye muligheter til å organisere og publisere forskningsmateriale, i tillegg til å gjøre dette tilgjengelig for andre forskere og samfunnet generelt (Organisation for Economic Co-operation and Development, 2015, s. 9).

3.2. EU og åpen forskning

Åpen forskning er en del av EUs visjon for hvordan forskningen i Europa skal utvikles. Det er en politisk forventning om at åpen forskning og innovasjon kan bidra til verdiskaping, bærekraft og internasjonalt samarbeid. Europakommisjonen ser på åpen vitenskap som et viktig ledd i sitt arbeide med å skape et «indre digitalt marked» som åpner digitale muligheter for samfunn og næringsliv, og som skal være med å gi Europa en ledende posisjon innfor den digitale økonomien (European Commission, 2016, s. 5).

Open innovation, Open science and Open to the World er rådende i EUs overordnede strategi (European Commission 2016), for hvordan forskning skal styrkes i Europa. EU-kommisjonens ambisjon er å lede utviklingen fremover, for eksempel gjennom rammeprogrammet til EU og gjennom policyutvikling. Prosjekter som blir finansiert gjennom rammeprogrammet til EU, har krav om at alle fagfellevurderte publikasjoner skal publiseres åpent tilgjengelig, og at tilhørende forskningsdata gjøres tilgjengelig, dersom det ikke foreligger spesielle hindringer eller at spesielle hensyn må tas (European Commission, 2016).

Våren 2018 la EU-kommisjonen fram fornyede anbefalinger om tilgang til og bevaring av vitenskapelig informasjon. I de nye anbefalingene oppfordres medlemsstater til å utarbeide nasjonale retningslinjer og handlingsplaner på en rekke områder. Områdene er fri tilgang til vitenskapelige publikasjoner, forvaltning av forskningsdata herunder fri tilgang, bevaring og gjenbruk av vitenskapelig informasjon, infrastruktur for åpen vitenskap, ferdigheter og kompetanse, samt insentiver og belønninger (European Commission, 2018).

3.3. Norge – nasjonale prinsipper og retningslinjer

Norske myndigheter vil, sammen med internasjonale organisasjoner som OECD og EU fremme mer deling og arkivering av offentlig finansiert forskning. Regjeringen har som mål at alle norske vitenskapelige artikler finansiert av offentlige midler skal være åpent tilgjengelige, og har fastsatt retningslinjer og tiltak for åpen tilgang til vitenskapelige artikler. I 2018 sluttet Forskningsrådet seg til Plan S som er et internasjonalt initiativ for åpen tilgang til vitenskapelige publikasjoner (Forskningsrådet, 2019).

Åpen tilgang til offentlig finansierte forskningsresultater har vært på den nasjonale politiske dagsordenen i over ti år. Politikktutformingene knyttet til tilgjengeliggjøring av forskningsdata kom for alvor i gang i Norge da OECD i 2007 anbefalte prinsipper og retningslinjer for tilgang til offentlig finansierte forskningsdata. Gjennom Stortingets behandling av St. meld. nr. 20 (2008-2009) «Klima for forskning», sluttet Norge seg til OECDs anbefalinger og retningslinjer (Forskningsrådet, 2019, s. 5). I Norge har forskningspolitikken dreid seg om at offentlig finansierte forskningsresultater skal komme samfunnet til gode. Forskningsrådet har siden år 2000 stilt krav om at forskningsdata skal arkiveres på en forsvarlig måte i minimum ti år. Det har også blitt utarbeidet en policy for tilgjengeliggjøring av forskningsdata. I 2017 lanserte regjeringen *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata* (Kunnskapsdepartementet, 2017). Strategien baserer seg på tre grunnprinsipper for hvordan forskningsdata skal tilrettelegges og deles. Det første grunnprinsippet dreier seg om at «forskningsdata skal være så åpne som mulig, så lukkede som nødvendig». Det andre grunnprinsippet handler om at «forskningsdata bør håndteres og tilrettelegges slik at verdiene i dataene kan utnyttes best mulig». Det tredje prinsippet handler om at «beslutninger om arkivering og tilrettelegging av forskningsdata må tas i forskerfellesskapene». (Kunnskapsdepartementet, 2017, s. 22-23).

3.4. Tidligere forskning og litteraturgjennomgang

Det har blitt utført flere studier på forskeres holdninger og erfaringer til lagring og deling av forskningsdata. Studiene viser frem utfordringene og kompleksiteten ved ulike aspekter som virker inn på arbeidet med forskningsdata. Tekniske, sosiologiske og økonomiske aspekter fremheves som viktige i forhold til forskeres arbeid med forskningsdata.

Flere studier viser til at forskere anerkjenner fordelene med åpen tilgang til forskningsdata som er offentlig finansiert. Blant annet undersøkelsen «Insight into digital preservation of research output in Europe» utført av PARSE.Insight (Kuipers and van der Hoeven, 2009). Forskerne mente at bevaring av forskningsdata er viktig, fordi det kan være med å stimulere til fremskritt innen forskning, og at det gjør det mulig å validere og reanalysere eksisterende data.

Jeg har også sett nærmere på tidligere masteroppgave av Therese Skagen (2015), *Forskningsdata på Handelshøyskolen BI: hva er forskernes behov?* Der 59 forskere besvarte et elektronisk spørreskjema om deres datapraksis og hvordan deres forskningsinstitusjon skal forbedre arbeidsprosessene for forskningsdata. Forskerne var positive til å dele data med andre forskere, men i liten grad med allmennheten. Barrierer for å gjøre data åpent tilgjengelig var sensitive data, og en mulig reduksjon av egen fremtidig publisering.

En tidligere masteroppgave av Live Håndlykken Kvale (2012), *Data sharing in life sciences: a study of researches at the Norwegian University of Life Sciences*, undersøkte forskeres holdninger til deling av forskningsdata innen livsvitenskapene ved Universitetet for miljø og biovitenskap. 147 respondenter besvarte et elektronisk spørreskjema. Spørreskjemaet er i hovedsak basert på en undersøkelse utført av PARSE.Insight i 2009, og tilpasset en norsk kontekst. Hun fant blant annet ut at forskere som deler data, ser på open science prinsipper som den viktigste grunnen til å gjøre dette, men at data først og fremst deles mellom kolleger og forskningsgruppe. Holdningene til å tilgjengeliggjøre data avhenger av hvor i karriereløpet forskeren befinner seg. Forskere med mer enn 20 års fartstid, er de som i størst grad gjør dataene sine tilgjengelig for andre. En forutsetning for å dele data er enerett til første publisering og akkreditering ved gjenbruk. Hun finner at et fåtall av forskerne har kjennskap til metadatastandarder og betydningen av metadata.

Tenopir et al. (2011) utførte en spørreundersøkelse blant 1329 forskere, der 75% var svært positive til å dele data med andre, og 78% var villige til å dele noe av dataene sine uten noen restriksjoner. Respondentene indikerte også at de var interessert i å bruke datasett fra andre forskere, hvis dataene var lett tilgjengelige. Til tross for dette var det bare 46% som gjorde dataene sine tilgjengelig for andre via internett, 36% var enige i at dataene deres var lett tilgjengelige, og mindre enn 6% gjorde alle dataene sine tilgjengelige. Resultatene reiser viktige spørsmål rundt kompleksiteten ved datadeling og gjenbruk, men avdekker i liten grad omstendigheter eller forhold rundt datadeling og gjenbruk (Tenopir et al., 2011).

Tenopir et al. (2015) gjorde en oppfølgingsundersøkelse av tidligere nevnte spørreundersøkelse om forskeres villighet til å dele og gjenbruke data. Resultatene viste at det fantes en økende

positiv holdning til å dele forskningsdata, men også en større bevissthet om risikoer, for eksempel feiltolkninger på grunn av data som holder dårlig kvalitet. I tillegg kunne mangel på og kjennskap til etablerte og beskrivende metadatastandarder være til hinder for å gjenfinne data, men også for å gjøre forskningsdata tilgjengelig for andre. Unge forskere er mer positive til å dele, men gjør det likevel sjeldnere enn forskere med lengre fartstid. Hvis forskere skal dele data, er det også viktig med gode insentiver for dette. Det fantes visse forskjeller i holdninger til å dele data, avhengig av hvilken fagdisiplin forskerne tilhørte (Tenopir et al., 2015).

For at forskningsdata skal kunne gjenbrukes må de først deles. Borgman (2012) har gjort en undersøkelse når det gjelder deling av data og kommer frem til at det ikke er så vanlig, og at det dessuten varierer veldig mellom ulike fagdisipliner og forskningstradisjoner. Det kan være vanskelig å tolke forskningsdata utenfor sin opprinnelige kontekst, noe som kan hindre deling og gjenbruk mellom ulike fagdisipliner. I tillegg er det utfordringer rundt rettighetsproblematikk og finansiering. Deling av data kan skje på flere ulike måter, enten direkte mellom forskere eller gjennom arkiv/databaser eller egne nettsteder. Det kan dreie seg om både rådata og bearbejdede data. Borgman peker på at det er bedre kartlagt hvorfor forskere ikke deler data, mens det er mindre kjent hvorfor forskere velger å dele data og hvorfor de velger å gjenbruke data (Borgman 2012).

Igjennom hele litteraturen på forskeres holdninger til datadeling og datadelingspraksiser eller atferd er det en spenning mellom idealet og virkeligheten. En stor andel av forskere støtter idéen om åpne data, men det er langt færre som har delt sine egne data i virkeligheten.

Noen av barrierene som hindrer datadeling inkluderer utilstrekkelig tid, behovet for å publisere først og mangel på finansiering (Tenopir et al., 2011). Wallis et al. (2013) fant at forskerne generelt var villige til å dele dataene sine så lenge deres rettigheter som forskere og forfattere ble beskyttet. For eksempel ønsker forskerne å publisere dataene sine først, og at dataene skal tolkes og siteres korrekt. Dessuten at prosessen med å dele data skal være enkel og praktisk (Wallis et al., 2013). Det er mer sannsynlig at forskerne deler dataene sine hvis prosessene er standardiserte, enkle og de kan få hjelp (Van den Eynden et al., 2016).

Ifølge Yoon (2017) kan mangel på metadatastandardisering hindre forskerne i å dele dataene sine åpent. Å forbedre disse standardene kan øke tilliten med å gjenbruke datasett fra andre. Yoon har intervjuet 38 forskere som har valgt å gjenbruke data fra andre forskere. Gjenbruk av data refererer til sekundær bruk av data, som ikke er ment å brukes for det opprinnelige formålet, men for å studere andre problemstillinger. Å vurdere dataenes pålitelighet blir viktig dersom man skal gjenbruke data. Anskaffelse av «korrekte» data er viktig i all forskning, fordi data som er «uegnet», det vil si data som ikke egner seg for forskningsspørsmålet eller er av dårlig kvalitet, kan føre til forvrengte og upålitelige resultater. Tillit viser seg å være en avgjørende faktor når en forsker vurderer å gjenbruke data generert av andre. Yoon konstaterer at tillit ikke er en statisk tilstand, men en dynamisk prosess, der tillit formes gjennom erfaringer med dataene i ulike steg. Tilliten kan forbli den samme gjennom hele prosessen, men den kan også gå tapt, eller gjenopprettes (Yoon 2017).

Å bestemme passende type og riktig mengde kontekstinformasjon for at skal kunne gjenbruke dataene er vanskelig. Kontekstinformasjonen som andre trenger for å kunne gjenbruke data er ofte vanskelig å artikulere eller utover forskernes opprinnelige forskningsformål (Faniel & Jacobsen, 2010). Ifølge Faniel & Jacobsen er det tre spørsmål forskere stiller seg når de vurderer å gjenbruke data fra andre: 1) er dataene relevante, 2) kan dataene forstås, og 3) er dataene pålitelige (Faniel & Jacobsen, 2010). Flere studier har funnet ut at metadata alene ofte ikke gir tilstrekkelig kontekstinformasjon til å kunne forstå dataene godt nok (Birnholtz & Bietz, 2003). Som medlemmer av et praksisfelleskap (Lave & Wenger), deler forskerne en forståelse for om forskningen er adekvat eller hensiktsmessig utført og de bruker denne forståelsen til å forstå andre kollegers data (Faniel & Jacobsen, 2010).

Tidligere forskning har vist at datapraksiser varierer på tvers av ulike underdisipliner og fagområder (Tenopir et al., 2015., Borgman 2012). Det er flere studier som har sett på datadeling med utgangspunkt i forskeres ulike fagdisipliner. Et eksempel er en rapport utarbeidet av Research Information Network (2008), i Storbritannia. Denne viser til at det ikke bare er variasjoner i ulike typer data, filformater, og «langtidslevedyktighet» i forskningsdata fra ulike disipliner, men også når det gjelder datapraksiser og kultur for deling av forskningsdata. I astronomi, genomikk og klassiske fag (humaniora/språkstudier) er det en

tradisjon for å dele data, og det er utviklet en infrastruktur som gjør det lettere å dele data. Innenfor andre områder var det mindre vanlig (Swan & Brown, 2008).

Til slutt vil jeg ta med noen studier som har vært spesielt nyttige å se nærmere på i forhold til min undersøkelse.

Zimmerman har utført en studie blant forskere innenfor feltet økologi. Økologi er studie av relasjoner mellom jordens organismer og deres miljø. Økologiske data er svært komplekse. Datasett har en tendens til å være små og svært mangfoldige, og metodene og teknikkene som brukes for å samle inn data varierte stort. Data ble lagret på lokale filsystemer, og metoder brukt for å dokumentere og forvalte data var svært forskjellige. Økologer hadde heller ikke en etablert infrastruktur for deling av data. Generelt foregår deling av data i hovedsak mellom nære medarbeidere og baserte seg på stor grad av sosial interaksjon (Zimmerman, 2003).

Hesse et al. så nærmere på hvordan forskning utføres innen fysisk oseanografi. For eksempel utføres forskning innen fysisk oseanografi ved hjelp av store forskningsfartøyer som har kostbart datainnsamlingsutstyr. Mange forskningsspørsmål krever data fra avsidesliggende steder, og er derfor avhengig av storskalaprojekter som involverer forskere fra mange land. Oseanografi er en empirisk data-tett vitenskap. Store databaser er vanlig. Allikevel er det også vanlig med små datasett som håndteres lokalt, til tross for store internasjonale datasentre (Hesse et al., 1993).

I en doktorgradsavhandling gjorde Winnie Tam (2016) en sammenligning av to underdisipliner innen geografi: fysisk geografi og samfunnsgeografi. Hun undersøkte blant annet hvordan data former og formes av disiplinære kulturer. De to underdisiplinene av geografi er ulike når det kommer til hvilke forskningsmetoder de benytter, hvor ofte forskerne samarbeider, samarbeidskonstellasjonene er av ulik størrelse og de samarbeider av ulike årsaker, data deles også på ulik måte, og av ulike grunner. Selv om denne undersøkelsen kun berører forskningsområdet geografi, er den interessant i et større perspektiv innenfor andre disipliner.

Studien viser også at forskningsmetode og egenskapene til dataene som genereres og anvendes har en sterk sammenheng. Tam sammenlignet i tillegg forskjeller i egenskapen til de data som ble generert, hvilke former for samarbeide som var vanligst, motiv og barrierer for å dele data og hvilke erfaringer forskerne hadde med deling innenfor de to underdisiplinene.

Undersøkelsen viste at datadelingspraksiser mellom fysisk geografi og samfunnsgeografi var svært ulike. De fleste forskere innenfor fagområdet fysisk geografi hadde erfaring med å dele data utenfor egen forskergruppe. Uformell datadeling var vanligst, selv om noen også delte data formelt gjennom nasjonale dataarkiv eller gjennom datatidsskrifter. Blant forskere innen sosial geografi var det kun noen få som hadde delt data utenfor egen forskergruppe. Etiske spørsmål rundt anonymitets- og konfidensialitet var noe de ofte måtte ta stilling til ved en eventuell datadeling, og noe som hindret forskerne i å dele data med andre. Tam konkluderte med at kvalitative data som er personlige i natur fra «myke» fagdisipliner som sosial geografi ofte genererer flere barrierer for deling av data, enn kvantitative data fra «harde» fagdisipliner som fysisk geografi som har preg av en mer upersonlig natur.

4. Metode

I denne delen vil jeg presentere metoden jeg har valgt for å belyse min problemstilling, redegjøre for begrensinger i metodevalget, dessuten komme nærmere inn på hvordan datainnsamlingen og utvalget av informanter ble gjort. Jeg vil også reflektere over forskningsetikk og diskutere kvaliteten ved forskningsprosessen og resultatet. Til slutt vil det bli tatt med et kritisk blikk på min egen rolle.

4.1. Vitenskapsteoretisk perspektiv

Ontologiske spørsmål dreier seg om hvordan virkeligheten egentlig ser ut og de grunnleggende antagelser om verdens beskaffenhet (Johannessen et al., 2016). Epistemologiske antagelser handler om hvordan man oppnår kunnskap og hva som kan bli sett på som gyldig eller akseptabel kunnskap innenfor et felt (Bryman, 2008, s. 4, Creswell, 2013, s. 21). Ifølge Creswell former forskerens ontologiske og epistemologiske antagelser hva slags vitenskapelige perspektiver og forskningsmetoder som vil være relevante å bruke for å belyse fenomenet som skal studeres (Creswell, 2013, s. 19-20). Den innfallsvinkelen vi ser verden ut ifra har betydning for hva vi ser, og hvordan vi fortolker og forstår det vi ser. Dette masterprosjektet kan knyttes til en sosialkonstruktivistisk tradisjon.

Sosialkonstruktivisme brukes om et perspektiv i sosiologi og andre samfunnsfag. Der menneskers virkelighetsforståelse betraktes som noe som kontinuerlig formes av opplevelser man har og situasjoner man befinner seg i (Tjora, 2021). Ifølge Berger og Luckmann eksisterer samfunnet som både en objektiv og subjektiv virkelighet, de skriver om hvordan utviklingen av mennesket og samfunnet skjer gjennom et gjensidig påvirkningsforhold. (Berger & Luckmann, 2000, s. 76). Det vil si at mennesket skaper samfunnet, samtidig som det også blir formet av samfunnet. Selve kjernen i sosialkonstruktivismen, er at den sosiale virkeligheten er konstruert og gjenskapes gjennom handlinger og interaksjon mellom mennesker (Ringdal, 2013, s.43). Det påpekes at idéer, begreper, så vel som organisasjoner og institusjoner er sosiale konstruksjoner, selv om de ofte framstår som en objektiv virkelighet (Ringdal, 2013, s. 43). Ringdal påpeker hvordan sterk konstruktivisme kan stå i fare for å helle mot antirealisme og relativisme. Det viktigste er ikke hvordan verden virkelig er, men hvordan vi oppfatter den. Sannheter er relative

og kulturavhengige. I de mest ekstreme versjoner av relativisme, blir det meningsløst å teste teorier mot observasjoner av virkeligheten, siden konsekvensen er at en empirisk sannhet ikke finnes (Ringdal, 2013, s. 44).

Relativisme og tro på at all kunnskap er konstruksjon, medfører en risiko for å konkludere med at alle framstillinger dermed er likeverdige (Ringdal, 2008, s. 44). Innenfor mange samfunnsvitenskapelige disipliner er det derfor en mer moderat eller mildere variant av sosialkonstruktivismen som er utbredt. Denne avviser ikke eksistensen av reelle objekter, men er opptatt av betydningen av *meningene* som tillegges slike, samt at meninger om fenomener, har betydning for handlingen rettet mot fenomenet (Tjora, 2021). Dette er ikke det samme som å si at vitenskapelige fakta er konstruksjoner. Snarere handler det om å forstå hvordan idéer, begreper og fenomener blir til og formes i relasjon til en bestemt sosial virkelighet, og innenfor bestemte institusjonelle rammer.

Sosial konstruktivisme blir også ofte beskrevet som en form for «interpretivisme», bakgrunnen er at individer søker forståelse av samfunnet de lever og jobber i. De utvikler subjektive meninger ut ifra deres erfaringer – og meninger rettet mot visse objekter eller fenomener. Disse meningene er varierte og mangfoldige, noe som leder til at forskeren vil undersøke kompleksiteten i tilknytning til disse objektene eller fenomenene, og hva som er årsak eller danner utgangspunkt for ulike syn og perspektiver. Hensikten med forskningen er å anerkjenne deltakernes syn og perspektiver på situasjonen (Creswell, 2013, s. 24).

I denne studien forsøker jeg å se nærmere på relasjonen mellom forskeres forskningsdata og deres fagdisiplin eller fagområde. Ved å fokusere på disiplinen eller fagområdet til forskeren, legger man vekt på at den spesifikke konteksten forskeren opererer i er forskjellig, og at dette er med på å forme forskerens erfaringer og meninger om forskningsdata. Denne konteksten er viktig for å forstå forskeres ulike utgangspunkt. Ifølge Creswell vil individers subjektive meninger ofte formes både gjennom sosiale interaksjoner med andre, og gjennom de kulturelle normene som råder innenfor et individs miljø eller omgivelser (Creswell, 2013, s. 25).

4.2. Metodologiske overveielser

Både teoretisk rammeverk, problemstilling, samt forskningsspørsmål er viktige ved valg av metode. Det samfunnsvitenskapelige perspektivet er utgangspunkt for denne undersøkelsen.

Samfunnsvitenskapelig forskning skal gi ny innsikt i samfunnsmessige forhold og prosesser, og bidra til økt forståelse om den sosiale virkeligheten (Johannessen et al., 2016, s. 25).

Samfunnsvitenskapens studiefelt er svært komplekst, og dreier seg om mennesker. Det handler om et mangfold av mennesker som har meninger og oppfatninger om seg selv og andre, og som tolker sine omgivelser og andre mennesker på ulike måter, meningene og oppfatningene er ikke stabile og vil stadig være i endring (Johannessen et al., 2016, s. 27).

Det er vanlig å skille mellom to hovedtyper av metoder, kvalitative og kvantitative. I kvantitative undersøkelser ønsker man å telle opp fenomener og å kartlegge utbredelse, i motsetning til kvalitativ metode hvor målet er å undersøke fenomener man ønsker en fyldigere forståelse av (Johannessen et al., 2016, s. 28). I en kvantitativ forskningsstrategi samler man inn informasjon som lar seg tallfeste eller utrykke i form av tall, mens det i en kvalitativ forskningsstrategi ikke samles inn tall på den måten, men heller informasjon som går i dybden. Med kvalitative metoder kan man fange opp personers mening og opplevelse som ikke lar seg tallfeste eller måle (Dalland, 2020, s. 54).

Det er fordeler og ulemper ved begge metodene. Kvantitative metoder gjør det mulig å innhente og systematisere informasjon fra store informantgrupper, samtidig kjennetegnes metoden av mer avstand til det fenomenet som studeres. Ved bruk av kvalitative metoder, som for eksempel intervju eller deltakende observasjon, får man større nærhet til studieobjektene, og dermed en rikere og dypere kunnskap om dem (Ringdal, 2018, s. 110-111).

Formålet med denne undersøkelsen er blant annet å få mer kunnskap om forskeres holdninger til deling av forskningsdata. Hva bunner disse holdningene i? Det er mange ulike forhold som virker inn på forskeres deling av forskningsdata. I og med at en masteroppgave er et tidsbegrenset prosjekt er det ikke mulig å undersøke alle sider ved slike forhold. Hensikten med studien er også å få en større forståelse for forskeres ulike utgangspunkt. Hva motiverer dem til å dele forskningsdata, og hva oppleves som hindringer for å gjøre forskningsdata tilgjengelig for

andre? Prosjektet har til hensikt å undersøke hvilke tanker og erfaringer forskere har rundt deling av forskningsdata. Jeg er interessert i å få vite mer om hvordan slik datadeling fungerer i praksis, og om det er noen spesifikke faktorer som har innflytelse på forskeres datadeling. Jeg har forsøkt å se nærmere på forskeres ulike utgangspunkt, ved å se på hva slags typer data forskerne genererer eller gjør bruk av, og å se dette i sammenheng med bakgrunn i deres ulike fagdisipliner eller fagområder.

4.3. Kvalitative intervjuer

Kvalitativ metode er spesielt hensiktsmessig når man ønsker å undersøke fenomener på en fyldigere måte, og kvalitative intervjuer er den mest brukte metoden for å samle inn data. En av fordelene ved kvalitative intervjuer er at det er mulig å få fyldige og detaljerte beskrivelser av det som studeres (Johannessen et al., 2016, s. 143). Det kvalitative intervjuet egner seg når man ønsker å forstå verden sett fra intervjupersonenes perspektiv. Intervju som metode blir ofte benyttet når man ønsker å forstå ulike aspekter av menneskelig erfaring (Kvale & Brinkmann, 2015, s. 135). Én-til-én intervjuer brukes gjerne når man ønsker å få innsikt i, og fyldige beskrivelser av informanters forståelse, erfaringer, oppfatninger, meninger, holdninger og refleksjoner knyttet til et fenomen (Johannessen et al., 2016, s. 144). Jeg vurderte derfor kvalitativ metode og intervju som den mest hensiktsmessige metoden. Ved å intervju forskere om konkrete utfordringer de har møtt og erfaringer de har gjort seg, håper jeg å bidra til mer innsikt i kompleksiteten rundt deling av forskningsdata. Jeg er interessert i å få vite mer om praksisen rundt datadeling.

Jeg har på bakgrunn av dette valgt å bruke semistrukturerte én-til-én intervjuer som metode. Et semistrukturert intervju har en overordnet intervjuguide som utgangspunkt, med en liste over temaer og generelle spørsmål som skal gjennomgås. Dette gir bedre muligheter for å kunne strukturere og sammenligne svarene, enn om man bruker en mer ustrukturert tilnærming som åpent intervju, der man ikke har en intervjuguide. Hvis man ikke benytter noen form for standardisering, kan det være vanskeligere å systematisere svarene i etterkant (Johannessen et al., 2016, s. 146). Intervjuguiden fungerte først og fremst som en støtte for meg under

samtalen, som en huskeliste over tema og spørsmål, jeg ønsket å komme inn på i løpet av samtalen. En semistrukturert tilnærming gir dessuten fleksibilitet nok, til at man kan følge opp interessante tråder og eventuelt inkludere andre spørsmål enn de som er listet opp i intervjuguiden. Samtidig vil intervjuet være avhengig av informantens svar og min oppfølging av disse. I noen intervjuer ble spørsmål utelatt dersom det ikke var hensiktsmessig å stille dem. Det var for eksempel ikke alle spørsmål som var aktuelle for alle informantene, noen hadde for eksempel ikke erfaring med å dele forskningsdata i arkiv/databaser eller erfaring med å gjenbruke forskningsdata fra andre forskere.

4.4. Utvalg av informanter og gjennomføring av intervjuer

Antallet informanter kan ikke være for stort siden datainnsamling og databearbeiding er en tidkrevende prosess. Kvalitative metoder kjennetegnes av at man forsøker å få mye informasjon av et begrenset antall personer. Samtidig må intervjumaterialet gi tilstrekkelig grunnlag for analyse. Hvor mange informanter som trengs, avhenger av forskningens formål og problemstilling. En tommelfingerregel er at utvalget skal være stort nok til at problemstillingen blir belyst. I tillegg snakker man gjerne om å oppnå et metningspunkt, det vil si at det bør gjennomføres intervjuer eller datainnsamling helt til forskeren ikke ser det som noen hensikt å intervju flere informanter, fordi de sannsynligvis ikke vil tilføre noen ny informasjon.

Johannessen et al. viser til at det ikke er uvanlig med et utvalg på 10-25 informanter, mens det i litt mindre prosjekter er vanlig med et utvalg på 10-15 informanter (Johannessen et al, 2016, s. 112). I dette masterprosjektet gjør blant annet begrenset tid at jeg tok høyde for at det var nok å intervju 10-15 informanter, jeg endte med 12 informanter. Dette kan imidlertid kritiseres for å være et for lite utvalg til å trekke slutninger, særlig fordi jeg har valgt å intervju forskere fra ulike fagdisipliner.

4.5. Strategisk utvalg

I kvalitativ forskning kan det benyttes et strategisk utvalg av informanter, det vil si at man velger deltakere som har egenskaper eller kvalifikasjoner som er relevante ut fra

problemstillingen og det vitenskapsteoretiske perspektivet som ligger til grunn for undersøkelsen (Johannessen et al., 2016).

Som case har jeg valgt å ta utgangspunkt i forskere som har tilknytning til polarforskning, siden det både fra enkelte forskere innenfor fagfeltet⁶ og fra Forskningsrådets side er nevnt at deling av forskningsdata har en ekstra stor betydning innenfor dette feltet (Forskningsrådet, 2020, s. 19). Slik polarforskning⁷ er definert, omfatter det all forskning som foregår innenfor et avgrenset geografisk område, og spenner således over mange ulike fagdisipliner (Aksnes & Rørstad, 2015). Å utføre forskning i polare strøk kan være spesielt kostbart på grunn av høye driftskostnader, til blant annet reiser, logistikk og drift av avansert utstyr og forskningsfartøy (Aksnes & Rørstad 2015, s. 44). Slik sett blir det viktig å dele på de forskningsdataene som er skaffet til veie. Mye av polarforskningen er dessuten offentlig finansiert (Aksnes & Rørstad 2015, s. 45), og det ligger som tidligere nevnt en oppfordring om å dele og tilgjengeliggjøre forskningsdata som er offentlig finansiert. Dessuten er det nevnt at tverrfaglighet er spesielt viktig innenfor polarforskning, og at det er en del samarbeid på tvers av fagområder (Forskningsrådet, 2020, s. 18). Jeg antok derfor at jeg ved å velge ut forskere som hadde en tilknytning til polarforskning, ville få tak i informanter som har erfaringer med å dele forskningsdata både utenfor egen forskningsgruppe og i dataarkiv/databaser eller lignende.

Når det gjaldt utvalget av informanter, forsøkte jeg å rekruttere informanter fra ulike fagdisipliner og fagområder. Det utøves polarforskning innenfor alle fagområder, men en stor del av majoriteten er naturvitenskapelig forskning (Aksnes & Rørstad 2015, s. 8). Jeg ville gjerne ha med forskere både med bakgrunn fra naturvitenskap, samfunnsvitenskap og humaniora. Beslutningen om å intervju forskere fra ulike fagdisipliner, og med ulike typer data, baserte seg på et ønske om å forsøke å få med ulike perspektiver på deling av forskningsdata. Er det noen åpenbare likheter eller forskjeller? Forskere fra ulike fagområder bruker gjerne ulike

⁶ Forskere oppfordres til delingskultur: https://en.uit.no/nyheter/artikkel?p_document_id=582263

⁷ Forskning som drives med grunnlag i materiale fra polarområdene (Arktis eller Antarktis), omkring fenomener med lokalisering i polarområdene, eller som tar direkte sikte på anvendelse i polarområdene (Aksnes & Rørstad, 2015, s. 13).

forskningsmetoder og genererer eller samler inn ulike datatyper. Hvordan vil for eksempel dataenes karakter virke inn på deling av forskningsdata? Ved å rekruttere forskere fra ulike fagområder kan man få et mer sammensatt bilde, på forskeres ulike perspektiver og utgangspunkt for deling av forskningsdata.

4.6. Rekruttering av informanter

Forskerne jeg rekrutterte til denne studien var tilknyttet ulike norske forskningsinstitusjoner, samtlige hadde deltatt i forskningsprosjekter innenfor polarforskning. Informantene er rekruttert via e-post, med noe oppfølging på telefon. Til sammen kontaktet jeg 26 forskere, alle inviterte fikk tilsendt et informasjonsskriv om prosjektet (se vedlegg 1). Av de 26 forskerne jeg kontaktet, var det tolv forskere som sa seg villige til å bli intervjuet.

Forskerne jeg rekrutterte til denne studien har bakgrunn fra ulike disipliner, og kommer fra fagområdene: naturfag, samfunnsfag og humaniora. Jeg rekrutterte åtte forskere med naturvitenskapelig bakgrunn, tre forskere med samfunnsvitenskapelig bakgrunn og en forsker innenfor humaniora. De forsker innen fysisk oseanografi, geovitenskap, statsvitenskap, marinbiologi, antropologi, toksikologi og arkeologi.

Det viste seg å bli en utfordring å få en balanse mellom de ulike fagområdene, siden jeg fant færre forskere å rekruttere innenfor samfunnsfag og humaniora. Selv om det utøves polarforskning innenfor alle fagområder, så er den store majoriteten naturvitenskapelig forskning. Innenfor samfunnsvitenskap var det i tillegg flere som takket nei til å delta, fordi deling av forskningsdata ikke ble ansett like aktuelt innenfor deres forskning, én jobbet for eksempel kun med analyse av dokumenter.

4.7. Intervjuguiden

I arbeidet med denne studien har jeg latt meg inspirere av Winnie Tams doktorgradsavhandling fra 2016, *Discipline and research data in Geography*. Jeg tok et utgangspunkt i hennes intervjuguide, men har tilpasset den til mitt masterprosjekt. Jeg har både fjernet, og lagt til noen spørsmål som jeg synes var relevant å få svar på i min studie. Intervjuguiden er delt inn i temaer som: forskningsfelt/fagdisiplin, datapraksis og dataenes karakter, datadeling generelt, deling av data i arkiv/databaser, datagjenbruk (se vedlegg 2).

Istedenfor å kun spørre forskere om deres holdninger til å dele data, har jeg også forsøkt å få en bedre bakgrunnsforståelse av forskeres ulike utgangspunkt for å kunne dele forskningsdata. Ved å forsøke å kartlegge noen karakteristika eller trekk ved fagdisiplinen de tilhører, hvilke forskningsmetoder de bruker, hva slags typer data de genererer eller gjør bruk av, og hva slags type samarbeidskonstellasjoner som er vanlig, forsøkte jeg å få en bedre forståelse av forskernes ulike utgangspunkt for å dele forskningsdata.

Jeg har i intervjuene med forskerne spurt om deres forskningsmetoder, og hva slags data som genereres eller innhentes i forbindelse med forskningen deres. Datapraksiser og forskningsmetoder er utvilsomt sammenvevd, siden forskningsmetoder i hovedsak handler om hvordan data blir samlet inn/generert og deretter analysert. Tam et al. mener det er sannsynlig at fagdisiplinenes egenart, som for eksempel forskningsmetode, forskningsspørsmål, prosesser og praksiser på tvers av og innenfor fag/underdisipliner, former egenskapene til data som blir skapt og hvordan data blir håndtert og lagret. Noe som igjen innvirker på mulighetene for datadeling og gjenbruk, og på hva som utgjør effektive dataforvaltningspraksiser (Tam, Fry & Probets, 2014, s. 723).

4.8. Pilot

Før jeg gikk i gang med intervjuene, gjennomførte jeg et pilotintervju med en forsker fra min egen arbeidsplass (IFE), for å teste ut intervjuguiden i praksis og for å se om noe behøvde å justeres. Slik kunne jeg få tilbakemeldinger på intervjuguiden, og jeg fikk blant annet forbedret spørsmålsstillingen, så det ble bedre flyt og en mer logisk oppbygning av intervjuet. I tillegg var det en fin anledning til å få trening i å intervju. Jeg fikk også en pekepinn på hvor lang tid intervjuene kunne ta. Denne forskeren hadde ikke noen tilknytning til polarforskning, og intervjuet og svarene herfra har ikke blitt innlemmet i studien.

4.9 Gjennomføring av intervjuene

På grunn av smittesituasjonen rundt Covid-19 våren 2021, ble det planlagt å gjennomføre intervjuene gjennom videoverktøyet Zoom. De som takket ja til å delta i undersøkelsen, fikk i tillegg tilsendt et informasjonsskriv om Zoom i forskningsintervjuer (vedlegg 3).

Elleve av intervjuene ble gjennomført via videoverktøyet Zoom, mens én av forskerne gjerne ville gjennomføre intervjuet over telefon, dette ble da etterkommet. Ved alle videointervjuene hadde jeg mitt eget kamera på, slik at informantene kunne se meg. Ti av informantene hadde også sitt kamera på, mens en valgte å ha det avslått.

I forhold til fysiske møter, vil det kunne være vanskeligere å avlese nyanser og reaksjoner over digitale videomøter. Men siden intervjuene ikke berørte personlige eller sensitive tema, mener jeg at det ikke hadde noen avgjørende betydning for utfallet. Selv om det nok er en fordel å kunne se personen man intervjuer, synes jeg de to intervjuene der jeg ikke kunne se informanten også fungerte fint. Jeg merket at jeg var veldig konsentrert om å lytte til stemmen, siden den var det eneste som kunne si noe om informantens sinnsstemning, men jeg synes ikke det hemmet kommunikasjonen og dialogen underveis i intervjuet. Allikevel kan jeg ikke helt utelukke at jeg hadde fanget opp andre nyanser, hvis jeg hadde møtt informantene fysisk.

For å kunne gjengi informasjonen mest mulig riktig, og for å unngå tap av informasjon, ønsket jeg å gjøre lydopptak av intervjuene. Elleve av informantene samtykket til at jeg kunne ta opp samtalen med lydopptaker. Opptaket skjedde ikke via Zoom, men via lydopptaker på mobilen. I forkant lastet jeg ned Nettskjema-diktafon appen, som er godkjent til bruk for opptak av forskningsintervjuer ved OsloMet. Filene blir kryptert, og kan ikke avlyttes før de er sendt til en personlig brukerkonto hos Nettskjema. Det er en egen passordinnlogging til brukerkontoen.

En av informantene ønsket ikke at det skulle gjøres opptak av intervjuet. Jeg tok da isteden notater underveis i løpet av intervjuet, og supplerte med å skrive ned flere detaljer etter hukommelsen rett etter at intervjuet var gjennomført. Det var en klar fordel å kunne ta opp intervjuene, jeg merket at jeg ble veldig opptatt av å få notert ned ting som ble sagt, og innimellom måtte jeg si ifra til informanten at nå må vi jeg stoppe litt, slik at jeg skulle rekke å notere ned det vesentligste. Det ble imidlertid vist forståelse for dette fra informantens side. Det var slik sett en klar fordel å kunne bruke lydopptak, siden jeg da kunne gi informanten min fulle oppmerksomhet.

Intervjuene ble gjennomført i løpet av juni og august 2021. Mange av forskerne hadde det svært travelt da jeg sendte ut de første e-postene i mai, en del var positivt innstilt til å bli

intervjuet, men spurte om jeg kunne komme tilbake til dem litt senere på sommeren. To intervjuer måtte dessuten utsettes til august. Derfor fikk jeg en god strekk i gjennomføringen av intervjuene. De tolv intervjuene hadde en varighet på mellom 43 minutter for det korteste og 1 time og 25 minutter for det lengste intervjuet. De fleste av intervjuene hadde en varighet på ca. 1 time.

4.10. Forskningsetikk

Ifølge Dalland handler forskningsetikk om å ivareta personvernet og sikre at de som deltar i studien, ikke blir påført skade eller unødvendige belastninger (Dalland, 2020, s. 168).

Siden jeg ønsket å gjøre opptak av intervjuene ble undersøkelsen meldt til NSD, for godkjenning av personvernopplysninger, bruk av taleopptak, informasjonsskriv med samtykkeerklæring, intervjuguide og informasjonsskriv om Zoom i forskningsintervjuer. Undersøkelsen ble godkjent før intervjuene startet (vedlegg 4). Ved rekruttering er det viktig at forespørsel rettes på en slik måte at frivilligheten ved deltakelse ivaretas. Det er viktig å informere om hensikten med prosjektet og hva opplysningene skal brukes til, at det er frivillig å delta, og at man kan trekke seg så lenge studien pågår (Dalland, 2020, s. 174).

Informantene ble også informert om at de kunne be om at notater og lydopptak ble slettet etter gjennomføring av intervjuet, dersom de skulle ønske det uten at det ville ha noen negative konsekvenser for dem. Alle intervjuer ble gjennomført med intervjupersonens informerte samtykke.

4.11. Konfidensialitet

Når det gjelder fortrolighet, eller konfidensialitet henspiller dette til en enighet som man inngår med informantene om hva som kan gjøres med dataene man samler inn, og som er et resultat av deres deltakelse. Dette innebærer som regel at private data som kan komme til å identifisere en person ikke avsløres (Kvale & Brinkmann, 2015, s. 106). I dette masterprosjektet skulle det være anonymt å delta.

For å verne om informantens konfidensialitet, vil ikke informantens navn oppgis i oppgaven, heller ikke ved hvilken institusjon de arbeider. Det finnes likevel en potensiell mulighet for at en

sammenstilling av variabler, kan avsløre identiteten til informantene indirekte, slik at man må gjøre en vurdering, også av andre opplysninger som potensielt kan være med på å identifisere en person. Jeg har for eksempel av den grunn valgt å ikke oppgi alder, kjønn eller stilling. Jo, flere variabler man sammenstiller, jo lettere vil det være å dra kjensel på enkeltinformanter, spesielt for personer i informantens nære fellesskap. Det ble imidlertid opplyst i informasjonsskrivet at det vil komme frem hvilken fagdisiplin informanten tilhører, hvilke forskningsmetoder som brukes og hva slags type data som genereres, siden dette vil være viktig for å kunne få en forståelse for konteksten når det gjelder deling av forskningsdata.

Som intervjuer er det mitt ansvar å reflektere over mulige konsekvenser for de personene som deltar i prosjektet. Dette prosjektet omhandler imidlertid ikke sensitive temaer eller personopplysninger, som kan ha noen større negative konsekvenser for deltakerne. Allikevel er det viktig å opptre varsomt. Man bør være klar over at den åpenheten og intimiteten som kan kjennetegne et intervju, noen ganger fører til at informanten gir informasjon som vedkommende senere angrep på at ble gitt (Kvale & Brinkmann, 2015, s. 107).

4.12. Bearbeiding av datamaterialet

Transkriberingsprosessen fungerer som en inngang til analysen. Intervjuene ble transkribert etter gjennomføring slik at det var mulig å analysere datamaterialet. Jeg leste først igjennom alle intervjuene, deretter startet jeg å markere viktige sitater og så etter ulike tema i tekstmaterialet. I analysen organiserte jeg data etter ulike kategorier, og kodet tekstsekvenser i Nvivo etter predefinerte koder. Deretter så jeg etter mønstre, sammenhenger og motsetninger i datamaterialet. Den kvalitative innholdsanalysen ble gjennomført ved gjentatt nærlesning av teksten av de transkriberte intervjuene.

4.13. Forskerens rolle

Et annet moment ved kvalitativ forskning er at den som forsker bruker seg selv som instrument (Johannessen et al., 2016, s. 230). Man snakker derfor om forskerens integritet som en avgjørende faktor for forskningens kvalitet, og de etiske beslutningene som treffes. Betydningen av forskerens integritet øker i forbindelse med intervju, fordi intervjueren selv er det viktigste redskapet til innhenting av kunnskap (Kvale & Brinkmann, 2015, s. 108).

Et rimelig kvalitetskrav er derfor at en foretar en selvkritisk vurdering av eget arbeid, og at en synliggjør valg som er tatt underveis. Funn bør presenteres så nøyaktig som mulig, og man bør reflektere over sin egen rolle som påvirker og fortolker, og etterstrebe en så fullstendig og nøytral undersøkelse som mulig. Det vil for eksempel si at man har et ansvar for å gjengi informantene på en sannferdig måte, og ikke tillegge dem meninger utover det som det finnes dekning for i dataene.

4.14. Refleksjon

Det subjektive elementet er til stede i all samfunnsvitenskap, men det gjelder i særlig grad kvalitative studier (Repstad, 2007, s. 138). I kvalitative intervjuer bruker forskeren seg selv som instrument. God gjennomgang av relevant litteratur var avgjørende for å bli trygg på tematikken og for å fange opp relevante innspill underveis i intervjuet i form av oppfølgings spørsmål. Jeg har både lest meg opp på forskningslitteratur og tidligere undersøkelser som gjelder forskeres deling av forskningsdata. I tillegg har jeg lest meg opp på policyer og retningslinjer som omhandler deling av forskningsdata.

Et intervju er en spesiell situasjon, og det kan ikke alltid antas at informantenes intensjoner om å dele data matcher deres fremtidige handlinger. Samtidig har jeg intervjuet forskerne om deres direkte erfaringer med å dele data og deres konkrete datadelingspraksiser, og derfor forsøkt å få et mer nøyaktig bilde av deres datadelingsatferd.

Interaksjonen mellom intervjuer og informant er også av en spesiell karakter siden temaet for samtalen er bestemt av meg som intervjuer, og som også er den som kontrollerer og styrer dialogen. Det var viktig for meg å skape en samtale basert på tillit og fortrolighet så mine informanter kunne føle seg trygge på å dele sine synspunkter, og slik gi ærlige svar. Det var derfor viktig for meg å understreke at jeg er opptatt av at forskere har ulike forutsetninger og utgangspunkt for å dele data, før jeg skulle i gang med å intervju de enkelte informantene.

5. Resultater og analyse

I intervjuene med informantene har jeg spurt om deres praksis eller erfaring med å dele data utenfor egen forskningsgruppe. Jeg har også forsøkt å se nærmere på forskernes ulike utgangspunkt og deretter se dette i sammenheng med deres fagdisiplin. Hvordan vil for eksempel dataenes karakter være med på å virke inn på deling av forskningsdata? Ved å rekruttere forskere fra ulike fagområder, har jeg forsøkt å få et mer sammensatt bilde på forskeres ulike perspektiver og utgangspunkt for deling av forskningsdata. Forskere fra ulike fagområder bruker gjerne ulike forskningsmetoder og genererer eller samler inn ulike typer data. Jeg har i intervjuene med forskerne derfor spurt om deres forskningsmetoder, og hva slags data som genereres eller innhentes i forbindelse med forskningen deres.

Ifølge Borgman kan disipliner være en praktisk, men langt fra feilfri eller nøyaktig måte å kategorisere ulikheter i datapraksiser. Siden grensene mellom fagdisipliner både kan være porøse og kunstige (Borgman, 2015, s. 56). Allikevel er det ofte slik at fag domineres av noen grunnleggende trekk. De kjennetegnes ofte av spesifikke studieobjekt, teorier, metoder og relevanskriterier som avgrenser det fra andre fag (Smeby, 2001).

I analysen vil jeg ta for meg forskernes praksiser og erfaringer med å dele forskningsdata generelt. I tillegg vil jeg se nærmere på forskernes holdninger til å dele forskningsdata utenfor sin egen forskningsgruppe.

Forskerne jeg rekrutterte til denne studien har bakgrunn fra ulike disipliner. Jeg rekrutterte åtte forskere med naturvitenskapelig bakgrunn, tre forskere med samfunnsvitenskapelig bakgrunn og en forsker innenfor humaniora. De forsker innen fysisk oseanografi, geovitenskap, marinbiologi, toksikologi, statsvitenskap, antropologi og arkeologi. Samtlige intervjupersoner var tilknyttet norske forskningsinstitusjoner. Kort presentasjon av informantene finnes vedlagt i tabell 1 på neste side.

Forsker	Fagdisiplin/fagfelt	Hoved-forskningsmetoder
N1	Fysisk oseanografi	Kvantitative - feltmålinger
N2	Geovitenskap/paleoceanografi	Kvantitative - geokjemiske analyser - sklerokronologi/"trering-kronologi"
S3	Statsvitenskap	Kvalitative - sekundærlitteratur/dokumentanalyse - intervjuer
S4	Statsvitenskap	Kvalitative - sekundærlitteratur/dokumentanalyse
N5	Fysisk oseanografi	Kvantitative -modellering/simuleringer -feltmålinger -laboratorie-eksperimenter
N6	Marinbiologi	Kvantitative -feltmålinger -biodiversitetskartlegging -kjemiske målinger/labanalyse
N7	Marinbiologi	Kvantitative -felteksperimenter -observasjonsteknologi
N8	Fysisk oseanografi	Kvantitative -feltmålinger -fjernmålinger/satellittobservasjoner -modelldata
N9	Fysisk oseanografi	Kvantitative -feltmålinger -fjernmålinger/satellittdata
S10	Sosialantropologi	Kvalitative -intervju -sekundærlitteratur/dokumentanalyse
N11	Toksikologi	Kvantitative -observasjoner/feltemålinger -kjemiske målinger -laboratorie-eksperimenter
H12	Arkeologi	Kvantitative og kvalitative -måledata (fra instrumenter) -visuelle observasjoner -feltarbeid

Tabell 1: N=naturvitenskap, S=samfunnsvitenskap, H=humaniora

5.1. Kort presentasjon av forskernes ulike fagdisipliner/fagfelt

5.1.1. Naturvitenskap

5.1.2. Geofysikk og geovitenskap

Felles for forskerne innen fysisk oseanografi og geovitenskap er at de virker innen forskningsfelt som har en sterk tradisjon og funksjonelle måter å dele data på. Ifølge Tenopir et al., kan forskere innenfor geofysikk sies å være i toppen når det kommer til datadeling, selv om det kan være variasjoner mellom ulike disipliner og underdisipliner (Tenopir et al, 2018, s. 891).

Oseanografer - kalles noen ganger havets meteorologer. Der meteorologer ser opp i atmosfæren på de fysiske tilstandene, ser oseanografer ned i havet, og prøver å vurdere hvordan det ser ut under vann. De forsker på fysiske forhold og fysiske prosesser i havet, og spesielt bevegelsene og egenskapene til havet. De ser for eksempel på vannstand, strøm, bølger, sjøis, havtemperatur og saltholdighet i havområder. Observasjonene brukes både til å utarbeide havvarsler, studere klimautviklingen og gi et grunnlag for forvaltning av havområdene. N8 forsker for eksempel «på å forstå prosesser og forstå hvorfor havet og klimaet oppfører seg som det gjør». Forsker N8 har blant annet vært med i forskningsprosjekter, der de prøver å utvikle et klimavarsel, for å kunne si noe om sannsynligheten for varmere eller kaldere tilstander over Norge, de neste fem årene.

Med denne forskningen kan de også si noe om de fysiske betingelsene for livet i havet. Ifølge forsker N5, så er det akkurat de samme ligningene og metodene i oseanografi, som i meteorologi. Oseanografisk observasjonsteknikk anvender i stadig større grad selvregistrerende instrumenter. I tillegg brukes det ulike former for fjernmåling som for eksempel satellitt. Satellitter samler informasjon døgnet rundt og sender den tilbake til jorden, men data kan også samles inn via mere manuelt betjente eller håndholdte instrumenter.

Hesse et al., påpeker at innenfor fysisk oseanografi samles det inn data fra avsidesliggende steder, noe som krever koordinering over lange avstander. Samarbeid og deling er nødvendig for å utføre fysisk oseanografisk forskning siden ingen enkeltpersoner og få institusjoner har råd til å utføre det på egen hånd (Hesse et al., 1993). Fysiske oseanografer trenger tilgang til store

databaser hvis de skal undersøke feltets forskningsspørsmål (Hesse et al., 1993, Zimmerman, 2003).

De forskningsmetodene som ble nevnt av de fire forskerne (N1, N5, N8 og N9) med bakgrunn i fysiske oseanografi var: observasjoner fra selvregistrerende instrumenter – blant annet fra satellitt, numeriske modeller, feltmålinger, og laboratorieeksperimenter. Oseanografiske instrumenter benyttes til å måle egenskaper og forhold i sjøvannet. Målingene kan enten foretas direkte eller for eksempel ved at man tar vannprøver, som siden analyseres.

N1: Vi er ute på båt, med jevne mellomrom, til dels så samler vi inn data mens man er på båten, vi tar prøver fra vannet, og sender ned instrumenter som logger der og da, og siden setter vi også ut instrumenter som står igjen til neste gang vi kommer og logger, og siden så er det i en viss utstrekning instrumenter som står ute og logger og som sender hjem data kontinuerlig via satellitt. Det er en stor investering bak hver datainnsamling både i tid, penger og infrastruktur.

Noen bruker også blant annet modelldata, og gjør laboratorieeksperimenter i tillegg.

N5: En del av det er en evaluering av klimamodellene, lage estimer av fremtidstilstanden i havet om ett år eller fem år frem i tid. [...] Det er jo tre grunnleggende ting man kan gjøre, man kan kjøre en generisk modell – simulere som klimamodellen, og så kan man dra ut i felt og måle ting ved hjelp av diverse instrumenter, og det tredje er at man gjør laboratorieeksperimenter. Jeg har egentlig gjort alle de tingene der, og du kan jo óg egentlig gjøre analytisk arbeid også, du bare sitter og regner og tenker på ligninger, men det har jeg gjort i veldig liten grad. Men ved de tre andre hovedgreiene er det jo datalagring involvert, og med klimamodellen er det veldig store datamengder.

I fysisk oseanografi er det et kjennetegn at det finnes en god del standardiserte og strukturerte data, fordi det er utviklet gode formater og standarder for fysiske data.

N5 Vi bruker jo en del sånne fancy filformater som netCDF, [...] når du begynner med modeller, så blir det veldig store datamengder – så det er en måte å komprimere dataene veldig bra på, altså det er bare en fil [...] men det ligger mange variabler [...] altså mye informasjon om hver enkelt variabel, for eksempel hvis du bare har en Excel-kolonne, og det bare står tall mellom 1 og 10, så vet du ikke hva det er. Om det er oppe eller nede, om det er temperatur eller dybde, eller istykkelse eller hva det er for noe, så i den netCDF-filen så ligger all denne informasjonen som sånne: langt navn, kort navn, enhet fra til i tid, fra til i rom, masse sånne predefinerte variabler som du skal fylle inn.

Og som på en måte ligger der helt eksplisitt da. Det er jo standardformatet på sånn geofysikk ting.

Ifølge forsker N9, så har man innenfor oseanografi etablerte standarder for lagring av data og også datadeling. Værvarsling og meteorologiske data er eksempel på data som karakteriseres som «enkle», og dette er data man holdt på med og standardisert i mer enn hundre år.

N9: Man kan jo begynne med det helt enkleste, de enkleste dataene som gjerne har vært samlet inn i mange år, der har man etablerte standarder, og eksempelet er da værvarsling og meteorologiske data, som har vært standardmålinger man har holdt på med gjerne i mer enn hundre år, og der har man etablerte rutiner. [...] Ja, innenfor oseanografi så har man etablerte, skal vi se en god del standarder for lagring av data, også datadeling.

Allikevel dukker det også opp en del nye data som ikke er standardisert og der det ikke finnes etablerte avtaler for hvordan disse skal lagres.

N9: [...] men så dukker det opp nye data, der man ikke har noen avtaler, det er masse nye data som dukker opp fordi at ny teknologi gir nye muligheter for å måle ting som vi ikke kunne måle for noen år siden, og der er det masse nye data som egentlig svever litt i cyberspace, det er ingen som har noen spesiell styring eller organisering av disse dataene, så det er stort spektrum av ulike måter disse dataene håndteres på og skal vi si gjøres tilgjengelig overfor brukere.

I enkelte databaser/datasentre kan forskerne få hjelp til å gjøre dataene sine tilgjengelig, men det er ikke nødvendigvis alle typer data de tar imot. Hvis det er litt mer unike eller eksperimentelle data, vet de ikke nødvendigvis hvordan de skal håndtere det.

N5: Men jeg har gjort mye eksperimentelle ting, og da er det jo jeg som måler og tar temperaturen og skriver det opp i notatboken min, og så lager jeg noen kolonner og så jeg har jo prøvd det og, da skulle jeg på en måte prøve å sende det til en database, men det var så originale data at de visste ikke hva de skulle gjøre med det.

Da blir det opp til forskeren selv å gjøre dataene tilgjengelig, og hvis det tar for mye tid eller blir for vanskelig, så er det ikke sikkert forskeren vil prioritere å bruke tid og ressurser på det.

N5: [...] du har en del standardformater, og så er det en del formater som ikke passer inn, og så skal du jo, hvis det blir for mye jobb å dele data, så er det grenser for hvor mye tid man har, så man kan jo finne på å droppe det, fordi det er for mye. Det er det største problemet.

Jeg intervjuet også en forsker innenfor geovitenskap, som hadde spesialisert seg innenfor paleoceanografi. Denne forskeren (N2) jobber blant annet med å rekonstruere klima bakover i tid. Bakgrunnen for dette er at de instrumentelle tidsseriemålingene som man har med for eksempel termometer er veldig korte, ettersom man bare har målt med termometer i ca. 100-150 år.

N2: Da vet man egentlig ikke nok om hva naturlig klimavariasjon er, fordi det er mange forskjellige dynamiske ting, ulike syklisiteter, som skjer på langtidsskalaen og det kan jo være multivokaliske, altså flere tiår og lengre. Ettersom man bare har målt med termometer i kanskje 100 -150 år, og de plassene er veldig få, særlig i havet, så må man ha andre måter å rekonstruere klima bakover i tid. For å kunne finne ut av - hva er klima egentlig? Hvordan fungerer det? Og da gjør jeg det for havet. Og det er en blanding av sediment, mikrofossil, masse forskjellige ting – kjemi. [...] En av de tingene jeg jobber med, er å rekonstruere klima ved hjelp av skjell, skjell man finner i havet - kuskjell, der har vi lånt alle metoder fra trering-kronologifolket, men det er folk som jobber på land, og vi jobber i havet, men det er fortsatt innenfor samme fagområde.

Forskeren forteller at for å kunne rekonstruere klima bakover i tid trengs det et «arkiv», som for eksempel sedimentkjerner, skjell, eller kalkalger, disse har et klimasignal innebygd i seg.

N2: Vi genererer geokjemiske analyser.. på ulike ting og tang som vi analyserer. Det kan også være aldersmodeller [...] der vi plotter tidsserier bakover i tid, det kan være tidsserier over temperaturer i havet, det er det ofte. I utgangspunktet er det slik at jeg studerer ulike arkiver, som sedimentkjerner eller skjell, eller kalkalger, eller noe i naturen som har et klimasignal bygd inn i seg, men det er forskjellige måter å få ut den signalen på, iblant er det geokjemiske analyser, iblant er det å måle tilveksten fra år til år, - for visse ting som skjell eller kalker har årlig tilvekst, så vi ekstraherer på en måte det vi kan ekstrahere ut ifra de her arkivene.

Ifølge forsker N2 er det innenfor denne forskerens fagfelt vanlig å sende inn dataene til arkivering i ulike store databaser, utenlands, – ett av dem heter PANGAEA (Data Publisher for Earth & Environmental Science), «og det er det man skal gjøre, man skal sende inn dataene til

de her databasene. Det er ikke alle som gjør det, men mye av dataene fins der av det jeg har publisert, men ikke alt.»

Forskeren opplever det å sende inn datasett til databaser som enkelt, og har ikke støtt på noen utfordringer, i forhold til det.

N2: Nei, i mitt felt er det enkelt, man lager stort sett bare en Excel-fil, og man får instruksjoner og det fins datafolk som jobber i de her databasene, og de kan man maile med fram og tilbake, og siden hjelper de til med å sammenstille det her, i det formatet som de trenger til databasen. Men generelt sett er det veldig enkelt, man har tidsseriene, man har sine data, og man har en liten beskrivelse, som de har gitt beskjed om at man skal ha, i mitt tilfelle er det breddegrad og lengdegrad og ja, man skal beskrive dataene, og det er lett. Det er lett for meg å gjøre, og det er lett for dem tror jeg også, og fikse dette her, i det endelige formatet som ligger ute på nettet.

Både innenfor geovitenskap og fysisk oseanografi sier forskerne at det er vanlig å dele data via databaser. Forskere innen fysisk oseanografi og forskeren innen geovitenskap nevner også at det er svært vanlig at tidsskrifter eller finansierer krever at dataene gjøres tilgjengelig ved publisering av artikler.

N5: Du er moralsk forpliktet og juridisk forpliktet til å dele, og publisere i åpne tidsskrifter. Og du skal ikke publisere noe sted hvor det ikke er åpent tilgjengelig. Publikasjonene også krever at data skal være tilgjengelig. Du får nesten ikke publisert uten å gjøre de mest sentrale dataene du har tilgjengelig.

N2: når man sender inn artikkelen, så skal man ha tenkt på hvor dataene skal ligge, datasett som man publiserer i artikkelen. Det kan - i hvertfall det som man har plottet, eller de tidsseriene som man publiserer i artikkelen, man kan også legge ved, noen form av rådata, for det kan godt være sånn at man har beregnet noe i artikkelen på grunnlag av et eller annet. Da kan man legge ved rådata, men det skal man ha et opplegg for, når man skriver artikkelen, og man kan gjerne når man submiterer artikkelen, skrive hvor disse dataene skal finnes tilgjengelig, og da senest. [...] For man får ingen penger om ikke dataene kommer til å være tilgjengelig, og det er jo litt likt, eller parallelt med at artikkelen man skriver også skal være tilgjengelig – open access. Man får ingen penger om man ikke har open access.

N1: I de tidsskriftene jeg publiserer i, så er det et krav om at man gjør data tilgjengelig etter publisasjon. Så det gjør jeg. Generelt er det et krav om å dele data i de tidsskriftene jeg publiserer i.

I tillegg virker disse forskerne å ha en sterk kultur for deling i utgangspunktet. N1 forteller at «om noen tar kontakt med meg og vil ha noe ut av mine data, så er det forventet at man deler dataene.»

N1: Ja, alle innen samme felt, bruker samme type instrumenter, vi gjør omtrent de samme tingene. Så det er veldig enkelt å bruke andres data i utgangspunktet. [...] Det er veldig vanlig på en måte. Ja, veldig vanlig. Det er typisk, altså jeg er med på en ekspedisjon til ett område ett år, så neste år er det en annen som er der, og tredje år er tredje person, så det er interessant å sammenligne alle de dataene. Så det er en høy grad av datadeling.

Dette viser at fysiske oseanografer er kjent og vant med en kultur, der data deles i stor grad. I tillegg jobber de ofte med beslektede fagdisipliner, der det også utveksles data.

N9: Vi bruker jo absolutt data fra beslektede fagdisipliner, innenfor marinforskning så har du flere, fysisk oseanografi og marinbiologi - og de må samarbeide tett og dele data, fordi de er avhengig av hverandre.

5.1.3. Biovitenskap - marinbiologi og toksikologi

De to forskerne jeg intervjuet innen marinbiologi sier begge at det nok er en kultur innenfor deres fagdisiplin at man i liten grad deler forskningsdata via databaser eller lignende. N6 kommenterer at folk har ikke vært vant til å tenke på den måten «de skal på en måte skrive artikkelen sin, og så skal de holde dataene tilgjengelig hvis noen spør etter dem, sånn at folk kan etterprøve det de har gjort, men den derre tanken om at nå har vi samlet inn data, nå må vi få det ut [...] det er veldig lite som det gjøres og det er vanskelig å snu den tradisjonen, fordi folk er vant til å stikke det i skuffen sin.» N7 mener imidlertid at det er noe som er i endring akkurat nå. «Ti år tilbake så ville det vært mye mer normalt og vanlig å ha dataene liggende på en harddisk med veldig liten innsats for å få de ut. [...] jeg vil si at det endres veldig i våre dager, og det bli mer og mer vanlig, og mye større aksept for at data er fritt tilgjengelig».

Den ene forskeren N6, jobber med alger og ser på mye på plankton, og variasjon gjennom våren. Forskeren ser for eksempel på hvilke betingelser som er viktige for utviklingen av algeoppblomstring i sjøen, og ser på sesongvariasjon på encellede organismer, i tillegg til å jobbe med akvaalger – som tang og tare, sett i forhold til diversitet og utbredelse. Forskeren

samler inn fysiske data som hentes inn fra et instrument, det kan for eksempel være sjøvannstemperatur, eller oksygennivå i havet. Likeledes gjøres det kjemiske målinger, der forskeren bruker vanninstrumenter på laben, for å gjøre målinger. I tillegg kan det være data som går på biodiversitet.

N6: Diversitet er det mye på, og så er det ganske mye, jeg bruker ofte DNA-metodikk til å identifisere organismer, siden jeg jobber med mikrober og encellede organismer, og da er det mye DNA-basert, det som heter metabarcoding, som er rådata.

Ifølge forskeren blir disse DNA-dataene, publisert i genbank, og de blir tilgjengelig for alle som måtte ønske å bruke dem. En klassisk database hvor man publiserer DNA-data, DNA-sekvenser eller tilsvarende data, er ifølge forskeren NCBI (National Center for Biotechnology Information). Andre typer data, som for eksempel fysiske og kjemiske målinger er ikke så vanlig å sende inn til databaser, så de blir gjerne liggende på forskerens egen harddisk. Forskeren forteller at fysiske parametere lagres som CSV-filer, måledata lagres som oftest i Excel-ark, mens DNA-dataene er så store at de må lagres i tekstformat, og da brukes SRS (Sequence Read Archive) for å publisere større mengder data.

Den andre forskeren N7, jobber med arktisk marinbiologi, effekt av klimaendringer og mørketidsbiologi. Forskningsmetoder som brukes er observasjonsteknologi og felteksperimenter. Det genereres veldig store data, siden forskeren jobber mye med observasjonsdata fra plattformer og sensorikk. Det samles inn fysiske og biologiske data, fra sjø og is. Denne forskeren sier at de «i størst mulig grad prøver å legge alt av data ut», og gjøre dataene offentlig tilgjengelige i en nasjonal database for forskningsdata.

Begge forskerne innen marinbiologi, sier at de ofte samarbeider med fysiske oseanografer, og kommenterer at det er stor forskjell på hva som kreves når det gjelder å håndtere ulike typer data, og at biologiske data ikke alltid er like lett å dokumentere.

N7: Enkelte fysiske oseanografiske data er nok lettere å arkivere, enn en del biologiske data. Data som tradisjonelt sett brukes i store globale modelleringsverktøy, de er som regel enkle og mer standardiserte enn mye av de biologiske dataene.

De største utfordringene med noen av de biologiske dataene er ifølge N7 «fravær, eller at det er for mange såkalte standarder, og at standardene veldig ofte er veldig godt tilpasset enkelte datatyper, men kanskje ikke i så stor grad en del andre datatyper. Kanskje spesielt akustiske data kan være kronglete å standardisere og gjøre godt nok tilgjengelig».

I tillegg kommenterer N6 at mange ikke er vant til å gjøre data tilgjengelig, slik at en del forskere heller ikke nødvendigvis har den kompetansen som skal til for å gjøre dem tilgjengelige.

N6: De enkleste er fysiske data, for der finnes det gode formater, sånn derre netCDF og CSV-format, som brukes mye til det, og det er det samme som meteorologer og værdata og sånn, og det har de holdt på med og standardisert i, jeg vet ikke hvor mange år allerede, [...], så der er det veldig gode systemer. Når det gjelder biologiske data så er det av og til litt mer tricky [...] Jeg opplever vel, jeg tror det er ganske mange typer data som kan være vanskelig å publisere fordi at det har vært en så dårlig tradisjon på å publisere datasett og dokumentere dem, og publisere dem med metadata og hele greia, sånn at folk er ikke vant til å gjøre det. Så da er prosessene fra å faktisk få det publisert som et datasett litt lang, for man vet ikke helt hvor man skal begynne. Det er jo kjempeviktig at man utvikler gode prosedyrer på dette, sånn at det er enkelt for den enkelte forsker å få ut dataene sine, det er ingen som har tid og krefter til å ta tak i det på egenhånd.

En annen forsker (N11) jeg intervjuet jobber innen toksikologi (miljøtoksikologi), og da innenfor to retninger. Den ene er hvordan miljøgifter forflytter seg i økosystemet og samler seg opp i ulike dyr i næringsnett, og den andre er å se på hvordan miljøgifter (forurensning) påvirker dyr, samtidig som at dyrene utsettes for andre ulike stressorer, som for eksempel klimaendringer eller næringsmangel. Metodene som brukes er gjerne observasjoner, i tillegg til at de gjør mye eksperimenter, hvor de registrerer ulike påvirkninger og hva slags effekter det har. Det er ofte snakk om små dyr, som spretthaler, hoppekreps eller humler. Hvis de er ute i felt kan de ta prøver fra fisk eller fugl. Forskningen genererer ulike typer data, og strekker seg fra ulike fysiske prøver til for eksempel kjemianalyser og statistikk. Forskningen krever også mye instrumentering.

N11: [...] bare for å gjøre de forsøkene, så trenger vi for eksempel klimaskap for å regulere klimaforholdene. Vi trenger avanserte lupr hvor vi kan følge dyrene og ta bilde av dem. Vi trenger dataprogrammer for å tolke bildene, så det er ganske mye, i tillegg til

statistikkprogrammer og slike ting». [...] «Det er ganske mange typer data vi jobber med. Og akkurat det med sånn DMP (datamanagement plan), er noe vi tenker mye på, og snakker om, og prøver å få litt kontroll på, men det er faktisk veldig vanskelig, selv innenfor en lab som vår. [...] Vi har jo mange masterstudenter, som skriver ned sine resultater i sin lab-journal, og som sliter veldig med å få dem inn elektronisk. Selv om vi terper mye på det, så er det et stykke fra hva vi prøver å få til, - til hva som faktisk blir gjort. [...] Det som jeg passer veldig på er at masterstudentene, når de er ferdige med masteroppgaven, at de har dataene bakerst i oppgaven. Det er en måte å få dataene ut fra deres skuff, og samle dem slik at vi har dem for ettertiden.

Forskeren forteller at

N11: vi har felles drivere hvor vi legger dataene, og vi prøver å ha det på en litt systematisk måte. [...] Vi prøver så godt vi kan, men jeg synes det er ganske kaotisk og uoversiktlig, faktisk. Vi bruker åpen driver for da får alle tilgang på alt. Så vi har en stor tillit innad i gruppa.

Dataene fra denne forskningen blir lagret internt ved institusjonen, men det er ikke vanlig å sende inn dataene til databaser eller dataarkiv. Når de publiserer artikler er det vanlig å skrive at data er tilgjengelig ved forespørsel, også fordi de ønsker å sikre at de kan ha en dialog med dem som skal bruke dataene, sånn at de brukes på riktig måte. Når jeg spør hva slags kultur det er rundt datadeling innen forskerens fagdisiplin, så sier forskeren «jeg synes at når vi skriver til hverandre så er det veldig positivt, da synes jeg de fleste er veldig positive, for da er det ofte at man oppretter et samarbeid. Men hvis man skriver og bare ber om dataene, så er det ikke så positiv respons». Dette kan tyde på at det ikke er like vanlig eller akseptert å dele data utenfor egen forskningsgruppe, dersom det ikke gjøres som en del av et samarbeid.

N11 forteller i tillegg at

N11: mye av det vi gjør er at vi bygger stein på stein, mye av det baserer seg på historikken og hva som har blitt gjort, så sånn sett bruker du kunnskapen til å forme nye hypoteser, men hvis vi trenger dataene direkte inn. Så vil vi nok kontakte de direkte. Men det er ikke så mye, vi har veldig mye aktivitet som vi har, så vi har på en måte ikke – det å koble opp mot gamle data. Det er ikke noe vi fokuserer veldig på.

Innenfor biovitenskap virker det ikke som tidsskriftene krever at dataene skal være tilgjengelig på samme måte som innenfor geovitenskap og fysisk oseanografi. Men forskeren N11 nevner at

«men det jeg ser blir mer og mer vanlig innen mitt fagfelt, er å publisere datasettene som en egen datapublikasjon. Jeg har ikke satt meg helt inn i det, men jeg ser det forekommer en del.»

Begge marinbiologene peker på at biologiske data kan være mer kompliserte å dokumentere, enn for eksempel mer standardiserte data som finnes innenfor oseanografi. Data som er vanskelig å dokumentere og arkivere, vil også være utfordrende å dele og gjøre tilgjengelig i databaser. Forsker N7 peker både på fravær av standarder og at det er for mange standarder, som en grunn til at det er vanskeligere å dokumentere biologiske data. (Det vil si at det ikke nødvendigvis er enighet om hvilke standarder som bør brukes, for de ulike typer av data). I tillegg kan det være egenskaper ved selve dataene, som gjør at noen biologiske data kan være vanskelig å dokumentere.

Borgman mener at ulike formål med å samle data, kan gi ulike grader eller spekter av spesifisitet – fra studier som beskriver spesifikke typer hendelser eller fenomen på den ene siden, til studier med den hensikten å simulere hele systemer eller modeller på den andre siden. Borgman viser til noen typiske forskjeller mellom data som genereres innenfor klimaforskning (som ulike typer værvarsler eller værphenomen), og mellom økologiske data. Værdata kan brukes til å beskrive eller forutsi temperatur, nedbør, vind også videre, og dataene kan kombineres med andre typer data og med prinsipper i fysikk, for å modellere klima fremover i tid (Borgman, 2010).

Ifølge Borgman utgjør økologiske data en kontrast eller en motsats til klimadata som brukes i modeller. Forskere som studerer et spesifikt fenomen som for eksempel oppblomstring av giftige alger, kan samle data i flere måneder eller år, for å fange hva som skjer før, i løpet av og etter en slik biologisk hendelse. Målet er å forstå prosessene som trigger en slik hendelse og hvordan disse prosessene utvikler seg. Det har vist seg å være vanskelig å aggregere forskningen fra slike komplekse biologiske fenomener til mere omfattende modeller som brukes i for eksempel klimamodellering, i hovedsak på grunn av ulikheter i dataenes egenskaper. I forskning hvor det studeres klima og fysiske prosesser, så finnes det oftere

etablerte fysiske konstanter⁸ og standardmålinger, mens forskning på biologiske organismer krever mer spesialiserte metoder og målinger, siden biologiske organismer kan være distinkt forskjellige fra hverandre (Borgman, 2010, s. 4).

Bowker hevder at økologiske data gir spesielle utfordringer når det gjelder dokumentasjon for sekundær bruk (Bowker, 2000). Zimmerman viser til at innenfor økologi så mangler det en tradisjon for deling, som delvis kan knyttes til disse spesielle utfordringene med å dokumentere data for sekundær bruk. Datasett tenderer til å være små og svært mangfoldige, og metoder og teknikker som brukes for å innhente og håndtere data varierer også stort. Denne variasjonen gjør det ofte vanskelig å beskrive økologiske data tilstrekkelig slik at andre kan gjenbruke dem (Zimmerman, 2003).

5.1.4. Samfunnsvitenskap - statsvitenskap og sosialantropologi

Innenfor samfunnsvitenskap har jeg intervjuet to statsvitere og en forsker innen sosialantropologi. De tre forskerne innen samfunnsvitenskap er veldig kvalitative, og bruker i liten grad kvantitative metoder. Alle tre bruker en del kildedata eller sekundærlitteratur. Kildedata kan gjerne være informasjon som er innhentet til andre formål enn forskning, som for eksempel offentlige dokumenter, selskapsinformasjon, informasjon fra media og policybeslutningsdokumenter, som for eksempel disse tre forskerne nevner som kilder de bruker. Men kildedata kan også være data fra annen forskning.

S3: På den ene siden er det jo innhenting av empiri, data gjennom litteratur, og nyhetskilder og offentlige dokumenter. Altså rett og slett datasøk på det som er tilgjengelig i offentligheten uansett, og en del av forskningen går ut på det, å sammenstille veldig mange forskjellige typer informasjon, for å skape et helhetsbilde. Så det er en dimensjon av det og som jeg nok vil si er ganske viktig, og da er jo på en måte dataene du henter inn den empirien, den informasjonen, den kilde ... Så det blir jo på en måte andrehåndsinformasjon.

S4: Jeg tilstreber å bruke mye, skal vi si originalkilder, og det vil jo da være offentlige dokumenter, men også veldig mye selskapsinformasjon, i det hele tatt veldig mye annen sekundærlitteratur og ikke minst media. Men det er liksom noe da som vi som kan

⁸ Fysisk konstant er en fysisk størrelse som antas å være både universell og ha en konstant verdi over tid. (Hofstad, 2020).

russisk, liker å gå så mye som mulig til originale russiske kilder, og ikke bare se på hva andre har skrevet om Russland.

Forsker S3 sier at det kan være vanskelig å dokumentere og arkivere de analysene man selv har gjort «hvis du har gjort en sånn generell empirisk datagjennomgang av andrehåndsinformasjon på nett og offentlige kilder og litteratur, så er det jo litt sånn, det er jo datainnsamling det og, men [...] Jeg har jo dataene, de blir ikke lagret en spesifikk plass, de blir ikke oppbevart heller slik at de er tilgjengelig for andre, selv om jeg vet at det også er et økende fokus på det, men det er jo noen begrensninger i temaet man jobber med».

Disse forskerne delte i hovedsak data kun innad i forskningsgruppen, deling av data utover egen forskningsgruppe er ikke vanlig, men for den ene statsviteren (S3) har det forekommet enkelte ganger at data har blitt delt på forespørsel ved privat utveksling. Blant annet et kvantitativt datasett, og dokument som ikke er offentlig tilgjengelig «hvis jeg har et dokument, eller hvis jeg har en slutterklæring fra et toppmøte eller et diplomatisk dokument, som jeg har fått tak igjennom kilder eller andre ting, så har jeg fått forespørsel om å dele det». Når jeg spør forsker S4 om vedkommende har fått forespørsel om å dele data, så er svaret «nei, jeg har ikke det».

S4 sier at «jeg forsker mest alene, men det er en del felles prosjekter og publikasjoner sammen med andre». S3 jobber også mest alene «det vil si ca. 75% alene og 25% i samarbeid med andre forskere. S3 forteller at «majoriteten av disse forskerne er innenfor samme fagområde, men i noen grad med forskere som jobber med tilstøtende emner, eller fra fagdisipliner som ikke er så langt unna, som jus og geografi». Forskeren sier at vedkommende jobber ganske uavhengig, - bestemmer i stor grad selv, og er lite avhengig av en forskningsgruppe.

Både S3 og S4 sier at de ikke har møtt på policy fra finansører eller lignende om å dele data. S3 sier likevel «det blir mer og mer vanlig også innenfor statsvitenskap, og det blir stilt krav til det [...], men det er ikke like enkelt eller relevant for alle typer forskning, det tror jeg ikke alle husker på».

Selv om det ikke er vanlig å dele data utenfor egen forskningsgruppe for disse statsviterne, har noen tidsskrifter inne statsvitenskap innført replikasjonsdatapolitikk, der artikler som er basert

på kvantitative data skal legge ut en kopi av datasettet på tidsskriftets hjemmesider.

Forfatteren skal da offentliggjøre en fil som oppgir de programspesifikke kommandoene som er nødvendig for å replisere funnene som er publisert i artikkelen. Begrunnelsen er kvalitetskontroll, bedre garanti mot juks, og at når noen andre kan se på dataene, så er det lettere at en gal konklusjon kan bli korrigert (Høyland & Hegre, 2008). Dette tyder på at det kan være ganske ulike datadelingspraksiser innenfor statsvitenskap.

Forskeren innen sosialantropologi (S10) jobber med klimatilpasning og omstilling til lavutslippssamfunnet i lokalsamfunn, i nordområdene. Denne forskeren ser på hvordan klimaendringen vil slå ut i lokalsamfunn, og prøver å koble menneske og miljø. Forskningen er veldig tverrfaglig, og vedkommende jobber med forskere fra ulike fagdisipliner «økologer, økonomer, samfunnsgeografer, miljøsosiologer, kanskje statsvitere, altså det spørres på inngangen, jeg jobber med jurister dersom jeg trenger det» og som oftest har prosjektene naturvitenskapelige deltakere. «Jeg jobber veldig tett på meteorologene og havforskningen, og alle de som har klimaforståelsen av klimasystemet».

Innenfor samfunnsvitenskap vil ofte informasjon om personer utgjøre en del av datagrunnlaget for forskningen. Forskeren innen sosialantropologi og den ene statsviteren nevner at de gjør intervjuer.

S3: Ja, så er det en del, så har jeg også gjort en del intervjuer, og gjør fortsatt en del intervjuer da, og samler direkte inn informasjon eller kunnskap fra intervjuobjekter som jobber med det temaet som man skriver om, og da er det jo lydopptak og transkribering.

S10: Det er mye semi-strukturerte åpne intervjuer, det er veldig mye jobb som går inn i en intervjuguide – og det er det den er, det er en guide, som er et dynamisk dokument, men som gir oss en retning på intervjuene og den bruker vi mye tid på. Så da har vi ganske lange enten individuelle intervjuer, gruppeintervjuer, fokusgrupper, litt ulike typer intervjuer, og så er det jo dokumentanalyser også, fordi det er mye sånne policybeslutningsdokumenter, som på en måte er styrende for hvordan ting er som de er.

Forsker S10 forteller at intervjudata anonymiseres og lagres.

S10: Vi har jo krav på oss for å arkivere de på et sikkert sted, og databehandling og sånne ting. Sånn at det gjøres, når det er vi som leder prosjektet så gjøres det hos oss.

Det er jo ofte de som gjerne leder prosjektet som har ansvaret, vi har jo rutiner for anonymisering av dataene, og hvis det transkriberes ute, så vet ikke den som transkriberer hvem det er, for da har vi ikke lagt inn noe navn, eller vi bare legger inn, eller gir de nummer. Så dataene vi må jo la de ligge er det ti år eller noe sånt noe.

Data fra intervjuer må behandles på en annen måte enn data som gir informasjon om verden ellers. Informasjon og data som kan knyttes til enkeltpersoner, vil komme inn under virkeområdet til personopplysningsloven. Data fra intervjuer er av etiske grunner gjerne konfidensielle. I utgangspunktet kan man derfor ikke dele slike data åpent.

S10: Nei, vi har ikke lov til å dele det, utover forskningsgruppen [...] men vi kan jo oppgi metadataene – hva slags type data det er, og hva det handler om» [...]. Men samtidig, vi som har med mennesker, det er veldig strenge etiske retningslinjer. Så vi kan ikke dele på samme vis, men vi kan dele noe av det.

S10 forteller at de kan dele metadata «vi kan jo oppgi metadataene, altså – hva slags type data det er – hva handler det om? Det kan vi legge inn i en database, og så vet folk at de dataene finnes».

I tillegg nevner S10 at de har publisert datasett, og da kan man bli kreditert.

S10: Fordi man kan få kredit på det, til og med vi gjør det. Hvis vi har dokumentanalyser som egentlig ikke kan publiseres som en artikkel, men som har verdi, så får de et sånt DOI-nummer. Og så publiserer du det som et datasett [...] da er det ikke noe analyse, altså for oss, så er det jo analyse, fordi vi har analysert dokumentene, men det er ikke noen konklusjoner, det er ikke noen diskusjoner, det er bare at du publiserer som et datasett. Og det er gjennom ulike kanaler man kan velge da, ettersom hvor det er interessant.

5.1.5. Humaniora – arkeologi

Humaniora forenes av studiet om menneskets kultur og dokumentasjon av menneskeheten, og drar veksler på alle slags tenkelige kilder. Samtidig kan humaniora ha fellestrekk med andre fagområder. Noen ganger regnes deler av noen fag innenfor for eksempel historie og arkeologi, som del av humaniora og mens andre regnes som del av samfunnsvitenskapen (Borgman, 2015).

Forskere innen humaniora er ofte avhengig av unikt materiale – det kan være brev, notater, fotografier og så videre, som kun eksisterer i fysisk form, En del av disse kildene er arkivert ved universitet, museum, offentlige etater og private institusjoner. Det å sammenligne objekter er ofte essensielt for forskningsspørsmål innenfor humaniora. Forskere kan samle inn informasjon på mange ulike måter, blant annet ved å ta detaljerte notater, ta fotografier, lage tegninger eller foreta 3D digitale skanninger. Digitalisering av objekter og deres representasjoner har transformert forskningsmetoder som støtter seg på slike sammenligninger (Borgman, 2015).

Jeg har intervjuet en forsker innen arkeologi. Arkeologi gir kunnskap om fortidens samfunn og innsikt i hva det er å være menneske. Ifølge Borgman kan det innenfor arkeologi benyttes kryssende områder av forskningsmetoder, og det kan trekkes vekslers på metoder, teorier og ekspertise fra både humaniora, samfunnsvitenskapen og naturvitenskapen (Borgman, 2015).

Forskeren (H12) jeg intervjuet innen arkeologi har «to bein å stå på». Den ene delen «handler om middelalderbyarkeologi og middelalderkeramikk, og den andre delen er bevaring av kulturminner der de ligger og klimaforandringer». Når det gjelder den førstnevnte – middelalderarkeologien, så vil det si «tolkninger av datidens hendelser, og helt konkret gjenstandsfunn og bygningsrester», mens når det gjelder den andre delen, så er det i stor grad innsamling av tverrvitenskapelige data». Nærmere bestemt innsamling av «måledata fra utstyr på arkeologiske lokaliteter – det kan være installert utstyr som måler temperatur, fuktinnhold og konduktivitet – altså ledningsevne og redox-parametere, som er parametere som tilsier om det foregår aktiv nedbrytning eller om det er stabile datamengder. Der kommer det inn data hver sjettede time, så det er utrolig store datamengder».

I den mer klassiske arkeologidelen er det gjerne mest kvalitative data som skapes. Det er tekstdata og refortolkninger. Mens delen som dreier seg om kulturminner og klimaforandringer, så samles det inn kvantitative data, disse benevnes av forskeren som «naturvitenskapelige data» eller «måledata». Forskningsmetoder som ble nevnt er feltarbeid og visuelle observasjoner.

Forskeren forteller at når det gjelder den mer klassiske arkeologidelen, så er man mer forsiktig med å dele forskningsdata, i alle fall før resultater fra forskningsprosjekt er publisert. Det er et

fåttall som deler data og et flertall som holder ting tett til kroppen. Forskeren forteller at arkeologifaget ikke er det mest delende, men viser til at det er viktig å dele grunndataene innen arkeologi slik at andre får mulighet til å omtolke og nytolke funnene. «Gjenstander, utgravningsplasser, tolkninger av ulike kontekster av sånt som er gravd ut, deles også noen ganger i form av rapporter. Og de skal egentlig være offentlige dokumenter, fordi med annen kunnskap kan man tolke ting på ulik måte.» Når man laster opp vitenskapelige artikler til ulike institusjonsarkiv, er det blitt mer vanlig å legge ved utgravningsrapporter, slik at også disse blir tilgjengelig for andre. «Og vi ser jo for eksempel, nå er sånne laserskanninger av landskap, sånne LIDAR-data de er blitt mye mer åpent tilgjengelig, tidligere var det kun forskerne som hadde betalt for å få foretatt skanningen som hadde tilgang til dataene.»

Forskeren påpeker også at når et felt er gravd ut, så finnes det ikke lenger, derfor blir dokumentasjon av grunndataene veldig viktig, slik at andre ved en senere anledning kan omtolke funnene.

H12: Ja, man måler ting inn, man fotograferer, man dokumenterer, fordi arkeologi er den eneste vitenskap som «dreper» informantene sine, når vi har gravd ut et sted så finnes det ikke lenger. Derfor er dokumentasjon og omtolkning utrolig viktig. Og dokumentasjonen er alfa og omega, fordi det betyr at da kan en annen arkeolog med større kunnskap, komme inn og nytolke de funnene, sånn at hvis man har trodd at man graver sånn og sånn, så viser det seg nei, det var ikke en brønn, det var en grop for tekstilfremstilling for eksempel, da skal man kunne finne ut av det ved å gå til grunndataene.

Forskeren forteller at på grunn av utviklingen med klimaforandringer, så er det kulturminner og klimaforandringene som nå har fått størst fokus. Utviklingen med klimaforandringene skjer altfor raskt, «og vi mister lokaliteter uten at vi har fått sjanse til å få undersøkt det». «Det er akkurat som å ta en bok du ikke har lest, så kaster du den, og det er det som skjer (...) fordi når en arkeolog graver seg ned igjennom et kulturminne så leser du kontekstene, det kan man ikke gjøre hvis et kulturminne er tatt i et jordras, eller er forsvunnet på sjøen (...) eller lag som var syv meter tjukke, plutselig har blitt komprimert til en halv meters tykkelse, så er det godt nok kanskje litt igjen, men det er utrolig vanskelig å tolke det».

Forskeren forteller at når det gjelder «klimaforskning og bevaringsstudier, så er det en helt annen åpenhet, og tradisjon for deling og samarbeid». Disse dataene har ifølge forskeren størst nytte når de deles med flest mulige, som kan bidra til tolkningen. «Jo flere som ser på dem desto bedre. Her gjelder det virkelig å dele, for å kunne tolke dem korrekt.» Måledataene eller de naturvitenskapelige dataene deles også via en egen nettside, der ganske mange personer fra ulike fag, og ulike universitet, ulike etater og private selskaper får tilgang til dem. Det kan hentes ut nyttig informasjon «for eksempel for en utbygger eller en kommune kan det være viktig å vite vannstand i det området, og når man skal planlegge ny infrastruktur trenger man å vite hvor dypt nye ledninger skal ligge for å ikke komme i konflikt med ting».

Forskeren jobber som oftest i tett samarbeid med andre forskere, gjerne fra andre fagfelt. «Vi jobber tett med geologer, hydrologer, altså folk som jobber både med jord og vann, fordi det er så viktig å ha den kunnskapen involvert. Og så jobber vi med kvartærgeologer, altså folk som jobber med veldig gamle jordlag, og paleoøkologer (...). Også selvfølgelig klimaforskere. I tillegg jobber vi med historikere og arkitekter, sånn som man har gjort under hele arkeologiens historie. Men den naturvitenskapelige delen øker bare hele tiden. [...] Det er ikke mulig å gjennomføre denne forskningen jeg driver med, uten at det må være tverrvitenskapelig.»

Når jeg spør om forskningsgruppens betydning, sier forskeren at «den er kjempeviktig». «Jeg kan ikke gjøre jobben min uten å ha den gruppen, fordi vi bidrar med ulik kunnskap, for å få fram den relevante kunnskapen om hva som kommer til å skje, og hva man bør gjøre. Og hvordan man kan avbøte effekten av klimaforandringer, når går grensen for at det man mister er så omfattende at det blir meningsløst.»

Forskeren forteller at når det gjelder den mer klassiske arkeologien, så er det mye kvalitative data. Tekstdata og refortolkninger. «Arkeologiske data kommer ikke med noen fasitliste, en annen person kan skape sine tolkninger med sitt kunnskapsgrunnlag, og det er sånn forskningen i arkeologi bygges videre på [...], fordi vi jobber med forhistorien, og ettersom tidsreiser ikke er mulig, så gjetter vi på – ut ifra de dataene vi har, hva som har skjedd før. Og det betyr at hver eneste nye utgravning, kan bringe kunnskap som snur opp ned på alt man har

trodd inntil da. Og det vil si at gamle data kan gjentolkes og omtolkes, og brukes til å fortelle nye historier.»

Jeg har spurt om forskeren har opplevd å ha behov for data som er frembragt av andre forskere som ikke er tilgjengelig?

H12: Ja. Det har vært slike grunnforskningsdata [...] det kan være at en forsker har publisert et kart med de tingene, han eller hun har tolket som noe, mens en annen forskeren har trengt å se hele planen med alle de ulike registrerte strukturene for å kunne omtolke og nytolke funnene. [...] Andre arkeologer både nåværende og fremtidige, trenger å ha hele planen med alle de ulike registrerte strukturene for å kunne, omtolke og nytolke funnene.» [...] Og det skal være tilgjengelig, men hvis det ikke er plassert i arkivet, men ligger under sofaen hjemme hos den arkeologen, så er det ikke tilgjengelig.

Spesielt for arkeologi er at en spesifikk gjenstand som regel blir avdekket og funnet bare en gang, men den kan bli gjenfortolket og omtolket gjentatte ganger. En annen karakteristikk er det store utvalget av datakilder og ekspertisen som kreves for å tolke dem. Bare ved å kjenne til konteksten til hvert enkelt objekt som blir studert kan de bli sammenlignet. Som et resultat av dette kan datakilder bli gjenbrukt gjentatte ganger, og ofte blir de mer verdifulle ettersom de akkumuleres over tid (Meld. St. 25 (2016–2017), Borgman, 2015).

5.2. Egenskaper ved dataene

5.2.1. Big data vs. «den lange halen av forskningsdata»

Betegnelsen «big data» eller stordata viser gjerne til samlinger av data som er så store og komplekse at det ikke er mulig å bruke tradisjonell dataprosessering for å gjøre bruk av dem.

N8: «Igjen det er ulike utfordringer med ulike typer data, med sånne klimamodelldata, så vil utfordringen ofte være størrelsen på dataene. Det er så mye data, og for hvert år som går blir disse dataene bare større. Og det er jo den utfordringen med sånne big data som folk snakker om. Hvordan skal du få tak i disse dataene, fordi det er liksom, det er ikke noe som er håndterbart lenger på din egen laptop.»

Dataene kan komme fra mange ulike kilder, men hvis de er standardiserte, kan de gi ny og viktig informasjon dersom de sammenstilles og analyseres. Data som befinner seg på den andre enden av skalaen, kan man for enkelhets skyld kalle smådata, selv om dette ikke alltid vil være en passende beskrivelse. Dette kan også være relativt store datasett, men et særtrekk er at de

ofte inneholder mer presis og detaljert informasjon enn stordata, og de vil også være mindre anvendelige til automatisk analyse og maskinell behandling (Forskningsrådet, 2021, s. 13).

Den ene forskeren N9 jobber spesielt mye med satellittfjernmålinger, data fra satellitt kan ofte karakteriseres som stordata. Forsker N9 sier at «siden jeg har holdt på med satellittdata, så har det vært en kolossal utvikling med bruk av satellitter til å måle og produsere data, og sørge for at dataene er tilgjengelige».

N9 «[...] hvis du tenker deg at det er en produksjonskjede for data, så har vi jobbet mest i begynnelsen av den kjeden, så etter hvert så skal dataene, legges inn i databaser. Så er det andre som kan bruke data ved å gå til databasene og så videre sånn. Så det er produksjonskjeden man jobber for å få til å bli bedre, eller den skal bli slik, at det er lettere for de som er lenger ut i kjeden - skal kunne gå og bruke data som er lenger opp i kjeden.»

Forsker N9 forteller at «det er en utfordring å få den prosessen til å bli bedre, og at data gjøres lettere tilgjengelig. Så det foregår en ganske stor innsats internasjonalt og nasjonalt for å gjøre data bedre tilgjengelig.» Satellittdata blir «håndtert gjennom internasjonale organisasjoner, og i Europa har vi ESA (European Space Agency) som er hovedaktør, og Norge er medlem i den organisasjonen, og har dermed full tilgang på å delta i programmene, og tilgang til dataene. Slik at når man snakker om satellittdata så er dette håndtert gjennom disse internasjonale organisasjonene, som da bruker ressurser på å tilrettelegge data og sørge for at de er tilgjengelige i ettertid».

Stordata genereres ofte innenfor fagområder med internasjonalt etablerte standarder, avanserte infrastrukturer og administrative støttesystemer, til forskjell fra smådata som ikke nødvendigvis har samme infrastruktur på plass (Forskningsrådet, 2021).

Stordata som engasjerer store forskerteam for å samles inn og håndteres, omtales noen ganger som «big science». «Big science» kjennetegnes gjerne av store team, langtidsprosjekter, og utstrakt bruk av instrumentering, data fra big science er gjerne store i volum, men vanligvis mer konsistent i strukturen (Borgman, 2015) Motsatt er data fra «small science», som gjerne omtales som «den lange halen» av forskning. Ifølge Wallis et al. tenderer disse dataene til å

være mindre i volum, lokal i karakter, og intendert for bruk i disse forskningsgruppene. Disse dataene er mindre tilbøyelige til å være strukturert på måter som tillater dem å overføres enkelt mellom forskningsgrupper og individer. De synes å utgjøre mesteparten av den vitenskapelige finansieringen og det er her hovedtyngden av forskning befinner seg. Ifølge Wallis et al. synes den store utfordringen å være og gjøre data fra «den lange halen» tilgjengelig og gjenbrukbar for andre, siden disse datasettene ofte ikke blir bevart (Wallis et al., 2013, s. 3). Smådata er ofte mer heterogene i natur og vil i mange tilfeller trenge individuell vurdering og tilpasning for at de skal kunne tilgjengeliggjøres for gjenbruk av andre (Forskningsrådet, 2021).

Datainnsamlingspraksis varierer veldig mellom «big science» og «small science» I «big science» tenderer ofte datainnsamling til å være godt planlagt og godt kuratert, dataene samles gjerne inn av høyautomatiserte instrumenter og dataene gjøres ofte gjenfinnbare (Wallis et al., 2013, s. 3).

N9 forteller at «på motsatt side har du da mer prosjektdefinerte datainnsamlinger som gjøres av noen få personer uten at det foregår gjennom en stor organisasjon, så det gjøres sånn individuelt forsker for forsker, eller i mindre forskningsgrupper og de dataene er da også i varierende grad håndtert i ettertid».

Borgman påpeker at i «big science» der det genereres store mengder data må man bli enige om felles formater og standarder, og det tilrettelegges for utviklingen av delte datainfrastrukturer, verktøy og tjenester. Disse dataene tenderer mot å være mer homogene i innhold og struktur. Mens jo mer av forskningsdataene som er «small science» i karakter, jo større er variasjonen på dataene, når det kommer til innhold, struktur og forekomster. De som forsker alene eller i mindre forskningsgrupper har større mulighet og fordelen av å kunne tilpasse forskningsmetoder, datainnsamling, instrumentering og analyse til forskningsspørsmålet, enn forskere som er avhengig av data fra automatiserte instrumenter, og massedigitalisering. Ulempen med denne fleksibiliteten er mangelen på forente standarder, og mangelen på kritisk masse til å utvikle bærekraftige dataressurser og delte felles ressurser (Borgman, 2015). Data fra «small science» eller fra «den lange halen av forskningsdata» vil man finne på tvers av alle fagdisipliner. Veldig ofte finnes disse bare på individuelle datamaskiner eller

institusjonsservere, og ofte med minimale eller ingen vedlagte metadata eller dokumentasjon, noe som utgjør et stort hinder for gjenbruk (Genova & Horstmann, 2016).

5.2.2. Kvalitative kontra kvantitative data

En dimensjon ved forskningsdata som det er relevant å se nærmere på, er hvorvidt dataene er innhentet ved bruk av kvantitativ eller kvalitativ metode. Hvilken metode som er brukt, kan ha innvirkning på hvor objektive dataene er, og hvorvidt det faktisk er mulig å skille «rene» data fra selve forskningsaktiviteten (Forskningsrådet, 2021, s. 12.) Prosessen i en kvalitativ studie preges av at beslutninger og tolkninger i stor grad hviler på forskerens egne skjønsmessige vurderinger, dette gjelder for både datainnsamling, bearbeiding, analyse og fortolkning (Repstad, 2007).

Becher og Trowler deler vitenskapsområdene inn i kategoriparet harde og myke fag, der humaniora og samfunnsvitenskap ses som myke og naturvitenskap som hard. En hard disiplin innebærer et fokus på universalitet, og forskningen er gjerne kvantitativt orientert. Ifølge Becher og Trowler karakteriseres harde fag av at de er kumulative, og nye funn bygger på tidligere forskningsresultater. Problemer kan ofte splittes opp i mindre enheter eller delproblemer, og løses hver for seg. En myk disiplin, på den andre siden, fokuserer i større grad på kvalitativ forskning. De er ofte preget av at samme forskningstema og grunnleggende problemer tas opp igjen av nye forskere og gis nye fortolkninger (Becher & Trowler, 2001).

Ved gjennomgang av intervjuene i denne studien kan det se ut til at forskere som bruker kvalitative metoder har flere barrierer mot å dele data, enn forskere som bruker kvantitative metoder. Innenfor samfunnsvitenskapen, vil ofte informasjon om personer utgjøre en del av datagrunnlaget for forskningen. Data fra for eksempel intervjuer tenderer mot å være relativt subjektive, spesielt sammenlignet med data fra naturvitenskap. Mens humaniora og samfunnsvitenskapene interesserer seg for mennesker og samspillet mellom dem, går naturvitenskap ut på å gi en objektiv forklaring på generelle fenomener og prosesser i naturen. Men det er viktig å påpeke at kvantitativ metode ikke nødvendigvis alltid fører til objektive data (Forskningsrådet, 2021).

Det er blitt stilt spørsmål ved gjennomførbarheten av sekundær analyse av kvalitative data. Gillies og Edwards undersøkte barrierene for gjenbruk, og fant ut at det er epistemologiske, metodiske, praktiske og etiske årsaker (2005). Roten til disse problemene ligger i den grunnleggende forskjellen mellom kvantitative og kvalitative data. Kvantitative data eksisterer gjerne uavhengig av forskeren, mens kvalitative data fra for eksempel intervjuer, kan være avhengig av interaksjoner mellom forsker og intervjuobjekter (Fry et al., 2008).

Begge forskerne innen statsvitenskap peker på at det ofte er en betydelig subjektivitet knyttet til samfunnsvitenskapelig forskning.

S4: [...] det er en betydelig grad av subjektivitet. Det er jo ofte det som egentlig kjennetegner mye av den type samfunnsforskning som vi driver. Vi ønsker selvfølgelig at det skal skape en viss - at det man konkluderer med skal være så tydelige som mulig, men det kan ikke stikkes under en stol at det vil være en form for subjektivitet, som man ikke kan utslette.

S3: Ja, det er jo de åpenbare utfordringer med reliabilitet, altså du må kunne stole på de dataene som er innhentet. [...] Det er samfunnsvitenskapelige temaer du opererer med, så du gjør ofte en vurdering, det er hele tiden avveininger og vurderinger, du tar hele tiden en kompleks virkelighet og reduserer, - deduserer den til en målbar eller brukbar kvantitet, og da det vil det alltid frafalle nyanser. Det er jo det som er skummelt ved å bruke andre sine data i samfunnsvitenskap. De nyansene kan av og til vise seg å være ganske verdifulle.

Ved bruk av kvalitativ metode er gjerne innhenting av dataene en mer integrert del av selve forskningsaktiviteten. Dette kan gjøre det vanskelig å skille tydelig mellom hva som utgjør dataene - det som analyseres eller fortolkes og hva som er forskerens intellektuelle bidrag. Slike prosesser er ofte hermeneutiske, hvor nye forståelser og nye elementer trekkes inn underveis (Forskningsrådet, 2021, s. 12). I samfunnsvitenskap, blir det ofte satt spørsmålstegn ved nyttheten eller relevansen ved å gjenbruke andres data, og dette har blant annet med omstendighetene og konteksten som ligger til grunn for dataene (Moore 2007).

Interessant nok, var det også noen av forskerne fra naturvitenskap som nevnte at data som kunne være vanskelig å dele eller arkivere, var subjektive eller kvalitative.

N2: Så er det jo mye data jeg egentlig aldri ville ha arkivert, og man kan jo spørre seg - hvor går grensen? For det er jo oftest en tolkning involvert i det datasettet som til syvende og sist blir publisert, og det gjelder fremfor alt de her kronologiene fra skjell, [...] for der har det jo vært en stor del av subjektiv tolkning. Der man bare by eye, må sitte og sammenligne masse - generere side opp og side ned med notater, det er ikke rådata som kommer ut av en maskin, og de kommer ikke til å bli arkivert.

N5: (...) og så har man jo hatt litt sånne kvalitative data, som man ikke har visst hvordan man skulle dele.

N11: Vi jobber mye med statistikk, vi ønsker å løfte bort fra det subjektive [...] hvis vi har kvalitative data, for eksempel, at vi observerer er det mye, lite eller medium av – jeg vil jo kalle det et kvalitativt mål – en vurdering – det er jo ofte mindre nøyaktig. [...] I den grad vi har data som vi må beskrive mer kvalitativt, så prøver vi å gjøre det på en systematisk måte, som gjør at vi kan tolke dem i etterkant. Vi er interessert i de generelle mønstrene, du har jo alltid noen sånne enkelt – det er alltid noe som oppfører seg annerledes, og det kan av og til være interessant, men stort sett prøver vi å se mer på – hva den generelle trenden er først og fremst.

Dette viser at det også innenfor naturvitenskap kan være en del beslutninger som tas basert på subjektiv tolkning, og som baserer seg på forskerens individuelle og skjønnsmessige erfaringer. Derfor er ikke naturvitenskapelige data heller nødvendigvis absolutt objektive. Forsker N11 beskriver kvalitative data som ofte mindre nøyaktige, sammenlignet med kvantitative data, og at de i større grad baserer seg på forskerens egen vurdering. Det kan også virke som om kvalitative data, basert på uttalelsene til forsker N2 og N11, er vanskeligere å systematisere, enn kvantitative data. Det tyder på at data som er mer utfordrende å systematisere og beskrive, og som i større grad baserer seg på forskernes eget skjønn, også kan være vanskeligere å dele på en effektiv måte.

Wallis et al. peker på at tradisjonelt er det svært strukturerte data som man er flinke til å dele og dra verdier ut av. Dette er typisk data man kan beskrive i form av definerte tabeller, og som er egnet for å lagres i databaser, og behandles i regneark eller lignende. Samtidig er dette den typen data det finnes minst av. Størsteparten av tilgjengelige data har ikke nødvendigvis en slik definert og fast struktur (Wallis et al., 2013).

S4 påpeker blant annet at formatet som de kvalitative dataene er i, kan gjøre dem vanskeligere å dele med andre.

[...] så er det egentlig dataenes karakter, de vil ofte ikke være samlet i tabeller som man kan kopiere over, det vil biter av forskjellige publikasjoner, eller ja kilder. Så derfor er ikke dataene liksom i et sånt format.

5.3. Holdninger og erfaringer til deling av forskningsdata utenfor egen forskningsgruppe.

Jeg har i min studie valgt å ha en vid tilnærming til begrepet datadeling. Borgman har definert deling av forskningsdata som «the release of research data for use by others» (Borgman, 2012, s. 1060). I denne studien fant jeg eksempler på at datadeling utenfor egen forskningsgruppe kan være å legge ved datasett til publiserte artikler, dele data i databaser, dele data via en personlig eller en institusjons nettside, dele data på forespørsel fra andre forskere. Noen av forskerne nevnte også datapublikasjoner, såkalte data reports. I min intervjuguide har jeg også tatt med spesifikke spørsmål som er knyttet direkte til deling av data i dataarkiv/databaser. Ifølge litteraturen på området er det innenfor en del fagdisipliner ikke vanlig å dele data via arkiv eller databaser. Jeg har derfor også intervjuet forskere om deres erfaringer med å dele forskningsdata utenfor egen forskningsgruppe generelt, siden ikke nødvendigvis alle forskere har erfaringer med å dele forskningsdata via arkiv.

En gjennomgang av intervjuene viste at forskerne generelt var positive til å dele data utenfor sin egen forskningsgruppe. Selv om forskerne fra samfunnsvitenskap også peker på at det ligger noen begrensninger i å dele det datamaterialet som blir til i deres forskning. Ved gjennomgang av intervjuene kommer det frem at spesielt forskerne innenfor geofysikk og geovitenskap har en tradisjon for og praksis med å dele forskningsdata, og at dette er noe som er forventet både av forskerne selv, men også av finansiører og tidsskrifter. Likevel viser det seg at også innenfor dette fagfeltet, der deling av forskningsdata er svært vanlig, så kan det være en del diskusjoner rundt detaljene ved deling av forskningsdata. N8 forteller for eksempel:

N8: Nesten alle vil jeg tro er veldig for at du skal dele data, at det skal være åpen tilgang, men jeg tror folk er veldig uenig likevel i detaljene i open access og open data. Altså gjelder det alle data, og skal det bli gjort med en gang? Så er det absolutt en diskusjon som har vært i det prosjektet, med som sagt litt mer unike data.

Nesten samtlige av informantene pekte også på noen betingelser eller forhold som må være på plass. Det som var nevnt flest ganger, og av nesten samtlige forskere var retten til å publisere sine forskningsresultater først, og at de må få tilstrekkelig med tid til å analysere dataene, før de skal tilgjengeliggjøres. Dette samsvarer med Wallis et al. sine funn som viser til at den vanligste betingelsen for datadeling er første rett til publisering (2013).

5.3.1. Førsteretten til publisering

Førsteretten til publisering handler i hovedsak om at forskerne som har samlet inn dataene, selv skal få tid til å utnytte dataene, før andre får tilgang til dem. Siden det bak ethvert forskningsprosjekt som regel ligger store investeringer i både tid og kostnader, er forskerne også bekymret for at noen skal ta eller kuppe dataene før man selv har fått sjansen til å publisere. Derfor har de fleste forskerne betenkeligheter med å dele forskningsdata, før man selv har fått publisert.

H12: Ja, altså når det kommer til tekstdelen og tolkning, og særlig når det er rundt ny banebrytende forskning, så har jeg lært det hardway at jeg skal være litt forsiktig med hvem man forteller allting til. Jeg fortsetter likevel å fortelle [...]. Dessuten er min forskningsfilosofi at man kommer lengst med å dele. Men det er noen personer jeg ikke forteller hva som helst til, fordi det er ikke alle som er like redelige i omgangen med andres forskningsresultater.

Forsker N1 forteller at tidligere så sendte vedkommende data til en stor database, der man kunne ha dataene passordbeskyttet, og når artikkelen var publisert tok man bare vekk passordbeskyttelsen. Mens nå skal dataene sendes til en annen database, og der har man ikke noen slik passordfunksjon «så da sender jeg ikke inn dataene før publikasjonen er ferdig. [...] Grunnen til det er at det er en så stor investering fra min side, at jeg vil ha første mulighet og rett til å publisere først».

Flere av informantene mener også man bør få en viss tid på seg, før dataene skal tilgjengeliggjøres for andre.

N8: Jeg ser ingen grunn til at forskningsdata ikke skal være offentlig tilgjengelig, med det lille forbeholdet om at jeg er veldig enig i at de forskerne som enten samler inn, eller lager og analyserer, eller setter opp slike eksperimenter, de skal få førsteretten til å

analysere dem. Da kan du si at det er en slags forsinkelse i den prosessen, om det er et par år eller om det er, altså det bør jo ikke være ti år, men ett år eller to, som gjør at de forskerne som har kall det betalt for disse målingene, eller har prosjekter som har betalt for disse målingene, og som har funnet opp disse eksperimentene, skal få lov til å komme med disse funnene. Så ikke bare alt blir lagt ut på nettet, så stikker folk av med det.

Flere av forskerne uttrykker bekymring for at dataene kan bli kuppet før de selv rekker å publisere. Det oppleves særlig sårbart dersom man har mer unike data eller data fra banebrytende forskning.

H12: Men det er noen ting - hvor man har jobbet med innsamling av data i årevis, så er det litt kjipt at noen finner dem, mens man sitter på slutten av å ferdigstille et manuskript med tolkning av de dataene. [...] Det er innmari gøy å se ens egne ting bli brukt i en helt annen kontekst, men før man har fått det ut på trykk så er det litt sårbart. Fordi da kan noen faktisk oppføre seg litt mindre pent. [...] Jeg har jo kolleger også som holder ting veldig tett til kroppen, inntil det er publisert, nettopp fordi det dessverre er noen som prøver seg med en lettvin: - «Nei, men nå er det vi som har publisert først». Men så skal den som har gjort hele jobben, men som får ut sin artikkel tre uker senere, skal da bli bedt om å henvise til den andre, fordi det var den andre som publiserte det her først. Nei, det kommer ikke til å skje, men vitenskapelig uredelighet er vel ikke noe man blir kvitt med det første.

N6 forteller at «utenfor gruppen, utenfor de nærmeste samarbeidspartnerne, så er folk litt usikre på om dataene er trygge eller om noen vil ta dem og publisere dem før deg og den type ting». Forsker N6 deler synet om at man skal ha førsteretten til å publisere data som en selv har samlet inn «det kan være jeg sier at her er dataene mine, du får gjerne se på dem, men du kan ikke bruke dem før de er publisert», men mener også at man burde kunne dele dataene sine åpent, raskere enn det som ofte gjøres i dag.

N6: Det er i alle fall en mye større andel av dataene som burde kunne deles åpent, mye tidligere enn man gjør, ikke alt, man kan ikke la alle andre få sitte og skrive artiklene dine. Man må jo få holde på noen sånne hovedpoenger. [...] Men det er nok mange som liker å holde på det lenger enn jeg synes er riktig. Selv om du ikke forutnyttet dem 110% fordi det ville tatt ti år, ikke sant, så kan man ikke sitte og holde på det så lenge, fordi at man gjør det med offentlige penger. Man må på en måte dele med hverandre, så hvis man ikke greier å utnytte det selv, så kan andre få muligheten. Så embargo-perioden kan ikke være for lang - til evig tid.

Forsker N7 svarer at vedkommende «støtter 110%» idéen om å dele forskningsdata åpent, og at «vi begynner nå å legge ut mer og mer i database X, og gjøre alt fritt tilgjengelig».

N7: Det er kun praktiske begrensninger», som gjør at ikke alle data gjøres tilgjengelig. Forskeren prøver å gjøre dataene tilgjengelig så raskt som mulig, men «i praksis så er det mest logisk å gjøre det i forbindelse med publisering. Så i praksis er det ofte knyttet opp til publisering, i teorien så prøver jeg å ikke tenke på den måten.

Men for de fleste er tidsdimensjonen for når man ønsker å dele data veldig viktig, at man får tid til å analysere og utnytte dataene selv først, og at man får første rett til å publisere. De fleste forskerne uttrykte at de ønsker å jobbe med dataene sine selv først, før de deler dem med andre. Altså at de får en viss «fredningstid», før dataene skal deles med andre. De viser til at de har lagt ned store investeringer i både tid og arbeid, derfor er det også rimelig å få utnyttet dataene først. De er ikke like enige i hvor lang denne embargo-tiden bør være. Enkelte mener at mange blir sittende for lenge på dataene sine før man deler dem.

En annen grunn til å vente med å dele data er av hensyn til stipendiater. N6 nevner at det er spesielt viktig for ph.d.-stipendiater at de får tid til å skrive publikasjonene sine, og at data derfor kan bli liggende noen år, innen de tilgjengeliggjøres.

Forsker N11 tenker at det på sikt er bra å dele data åpent, men synes også at det er viktig å gi forskerne den tiden de trenger for å jobbe med dataene

N11: Det er jo ikke sånn at tre måneder etter at dataene er ferdig så skal det være åpent, fordi det kan ta flere år å skrive vitenskapelige artikler. Så det er jo først og fremst viktig når du har stipendiater å beskytte dem, sånn at de for jobbet i fred og ro med sine ting. Selv etter at de er ferdig så er det sjelden at ting er publisert, så jeg synes det er viktig å gi miljøene ro til å publisere fem år etter prosjektets slutt eller noe sånt.

5.3.2. Forskningsdata – forskerens kapital

Når jeg spør forsker N2 om hva vedkommende tror er den største hindringen for datadeling blant forskere, svarer denne forskeren at det er veldig viktig å få ut så mange artikler av datasettet som mulig, siden man har investert så mye i form av både tid og penger.

N2: Man vil selv ha publikasjonen på datasettet, og ikke gi det til noen andre, forsøke å skvise ut så mange artikler som mulig fra det samme datasettet, og ikke da på en måte ha sittet og produsert data til noen andre, som på en måte enkelt da bare takk – så sitter de og skriver en artikkel på en måned. [...] hinderet er at man vil ha ut så mye som mulig ut ifra den jobben man har lagt ned, og spesielt i mitt felt, når man har dratt inn hver eneste krone selv, det er utrolig vanskelig å få forskningsprosjekter og da er det veldig, veldig viktig at man får ut så mye som mulig, i form av artikler, som man kan ha på sin cv, for å kunne få mer penger, eller ære og berømmelse.

Når man først har samlet inn et godt datasett, så vil det kunne gi grunnlag for nye søknader og mer forskning. Da kan det bli et disintensiv å skulle dele datasettet, i alle fall før man selv har fått utnyttet dataene selv.

En annen forsker uttrykte også en viss skepsis til idéen å skulle måtte dele data åpent, hvis man selv har tenkt å bruke dataene videre. Å dele dataene kan da føre til at noen andre gjør det som en selv har tenkt å gjøre, og hindre at man får publisert de artiklene man har tenkt seg eller utnyttet dataene maksimalt, fordi noen andre kan komme deg i forkjøpet.

S4: Sånn subjektivt, særlig hvis du har tenkt å bruke dataene selv videre. Kan du da sitte rolig å se på at dette gis bort til andre, og de gjør kanskje de tingene som du hadde tenkt å gjøre? [...] og andres bruk av det kan gå på bekostning av deg.

Selv om forskerne i hovedsak gir sin støtte til idéen om deling av forskningsdata i prinsippet, gir også noen forskere samtidig uttrykk for en ambivalens når det gjelder å skulle dele data åpent. På den ene siden ønsker de det, mens på den andre siden må man også tenke på hvilke konsekvenser det kan ha for en selv.

S3: Men andre typer betenkeligheter jeg har er [...] hvis jeg bruker veldig lang tid på det ene datasettet, skal andre få det gratis da, så har jeg brukt så mye tid med å jobbe med det [...] for den dataen er veldig verdifull for meg og det er en del av min jobb, så hvorfor skal jeg gi det til andre? Men samtidig har jeg en generell tilnærming til akademia, om at jeg tror ikke akademia er et nullspillsum, så dermed deler jeg heller litt for mye enn for lite rett og slett.

Sitter en på datamateriale, sitter en også på «kapital» for nye søknader og for mulige publikasjoner. De som har samlet inn data kan ha investert betydelig arbeidsinnsats og

kreativitet i planlegging og innsamling, og derfor kan de kjenne på et eierskap til dataene de selv har vært med å frembringe. Forskere kan derfor ofte uttrykke en ambivalens, på den ene siden støtter de idéen om åpen forskning og åpne data, på den andre siden må dette også balanseres med å ikke ofre egne karrieremuligheter og mulighet til å få tak i flere forskningsmidler for å finansiere nye forskningsprosjekter. Å dele data kan da samtidig være å gi avkall på eller risikere å miste noen av disse fordelene.

N2 forteller at «det fins selvfølgelig folk – og det er ikke uvanlig, som sitter og holder på en del datasett. Der man for eksempel har gjort en liten beregning basert på et datasett, der beregningen er publisert, men der selve det grunnleggende datasettet, ikke trenger å være det. Og det er ganske vanlig. [...] Folk sitter på datasett som de vil recycle og som de har planer å gjøre noe med.»

N2: Det er fortsatt sånn at alle datasett ikke ligger ute, folk legger ikke ut alt. Spesielt den gruppen som har tenkt å jobbe videre med dataene, eller som bare er imot hele opplegget, de gjør bare ikke det her. Det blir feil, men det blir færre og færre, fordi tidsskriftene krever det, finansieringskildene krever det også, så det blir færre motstandere, men for tjue år siden var det annerledes, da var det bare deler av alle datasett som lå ute.

5.3.3. Gjensidighet

I fagdisipliner innen naturvitenskap har det ofte vært etablert en etikk, kultur og tradisjon for datadeling, basert på uformelle og personlige kontakter, tillitsforhold og gjensidighet av datautveksling på individuell basis gjennom forskernettverk (Wallis et al., 2007, Zimmerman, 2008).

Når jeg spør en forsker N1, om hva vedkommende tenker om å dele data utenfor forskningsgruppen, forteller forskeren «at det er for det meste ok [...] jeg jobber gjerne internasjonalt, og det er få som jobber i det samme geografiske området, så jeg kjenner de aller fleste som gjør det».

N1: Jeg deler data med dem, men da samarbeider jeg med dem. Om noen tar kontakt og vil ha noe ut av mine data, så er det forventet at man deler data, det er et samarbeid rundt det.

Forskere har tradisjonelt delt data, dokumenter og andre intellektuelle ressurser med hverandre gjennom pålitelige relasjoner og forskernettverk. Data kan derfor utveksles eller byttes med andre data og ressurser, eller brukes for å inngå samarbeidsprosjekter. Hvis data deles åpent med alle, kan forskere miste muligheten til å bytte data privat, og dette kan skape et disinsentiv for deling (Tenopir, 2018).

N11: det har jo vært en enorm innsats i det å fremskaffe de dataene. Så jeg synes det er bedre å dele data, som en del av et samarbeid, hvor man er med på den videre bruken og tolkningen av dem. [...] Jeg har blitt kontaktet personlig og da har vi prøvd å finne ut, hva skal det brukes til og om vi kan få til et samarbeid på en måte på det, - det er gjerne disse kjemiresultatene. Jeg har også fått forespørsel om å dele koding, som vi bruker i statistikken.

Forskere kan også kjenne på forpliktelser overfor dem som har frembragt dataene og som de ønsker å bruke videre. Respekt for andres arbeid, kan også gjøre at de ikke bare ønsker å «forsyne seg» av dataene, uten at man oppretter en form for samarbeid. N11 forteller for eksempel at:

N11: Jeg har skrevet sammen data fra ulike innsjøer, da kunne jeg jo bare hentet dataene, men da har jeg heller kontaktet forfatterne, og spurt om de vil være med inn. Både fordi jeg synes det er en real ting å gjøre, men i tillegg har de bidratt veldig inn i skrivningen, og de kjenner dataene godt.

N2 forteller blant annet at «det er utrolig uvanlig at folk tar noen andres datasett og begynner å publisere en artikkel på egenhånd, uten at noen av disse folkene er med.»

Det tyder på at en form for samarbeid er vanlig dersom man skal gjenbruke andres data, og at deling av forskningsdata ofte innebærer en form for samarbeid og sosial interaksjon, mellom den eller de som skal gjenbruke dataene og den opprinnelige dataskaperen eller opphavelige forskningsgruppen. Ifølge Wallis et al., er slike «forhandlinger», gjerne vanlig for forskning som utføres «i den lange halen av forskning», det vil si for forskning som er utført av en enkeltforsker eller av mindre forskerteam (Wallis et al., 2013).

5.3.4. Tid til å kvalitetssikre dataene

I tillegg til førsteretten til å publisere sine forskningsresultater først, var det flere forskere som nevnte at de måtte få nok tid til å kvalitetssikre dataene før de gjøres tilgjengelig for andre. Når deling av data trekker ut i tid kan det også komme av at kvalitetssikringen er tidkrevende. N5 forteller at «det kan jo være man må luke ut feil og så videre, og så har du hatt masse å gjøre, og så går det et år eller to. Så er du på en måte ikke ferdig med å kvalitetssikre eller du skal kanskje sjekke mot en modell, et eller annet sånt før du stoler helt på variasjonene». Forsker N11 mener det er et voldsomt press på det publiseringsystemet, men at det er viktig å få tid til å kvalitetssikre dataene.

N11: Nå har vi akkurat hatt noen prosjekter hvor det har vært mye diskusjoner om hvordan skal dataene analyseres. Hva betyr de egentlig? Fordi det er viktig at det er enighet blant medforfatterne, før ting sendes inn. Og sånne ting vil ta ekstra tid. Hvis ting går helt greit, kan det publiseres kjempfort, men av og til kan det ta årevis. Som jeg sa så er jeg veldig for at det skal være en transparens fordi det ligger på en måte i bunn av forskningens natur, at det skal være etterprøvbart. Og så er det veldig mye offentlige penger inni det her, så det er jo ikke mine private data. Men det er viktig at jeg går god for kvaliteten både på resultatene og også tolkningen og bruken av dem. Så det er på en måte et ansvar som man må ha tid til å gå inn i på en ordentlig måte.

Forskeren N9 nevner «Climategate» som «sånn klassisk episode for noen år siden der noen jukset med klimadataene». Det kan slå tilbake på forskeren om dataene ikke er kvalitetssikret godt nok, siden det kan få store konsekvenser hvis dataene blir misbrukt. Derfor vil denne forskeren først legge ut data, når vedkommende er sikker på at kvaliteten er god nok. Hvis «dataene er 100% pålitelige, da er det greit, fordi du vet at de står seg.» Men forskeren forteller at så har du gjerne data du er usikker på, det kan være at dataene er samlet inn ved hjelp av et nytt instrument, og man vet ikke hvor nøyaktig eller godt det er, og derfor tar det tid å utvikle nye dataprodukter, for man må bygge opp tilliten til at dataene holder kvalitet.

N9: Jeg vil ikke bare legge de ut med en gang, uten å vite og ha tillit til hvor gode de er, for da er det igjen lett for at noen kan misbruke de dataene, og bruke de utenfor de forutsetningene som ligger der. Så det er dette med å etablere dataprodukter med kvalitet, du ønsker å gjøre dette før de legges ut offentlig. Det kan ofte ta lang tid, du kan ha samlet inn data for flere år siden, men du er ikke sikker på om de dataene er pålitelige eller gode nok til at de skal legges ut.

Forsker N9 sier det kan være litt problematisk, hvis du kommer i en situasjon der du ikke kan dokumentere kvaliteten og dataene blir misbrukt «så kan dette slå tilbake på deg selv, fordi du har produsert noe data, men du kan ikke gå god for kvaliteten. Så da har du et problem der». Forskere kan være skeptisk til å gjøre dataene tilgjengelig for andre før man er sikker på at dataene er kvalitetssikret. Forsker N9 forteller at dersom dataene man legger ut viser seg å ikke være pålitelige nok, og at dataene ikke måler det de har tenkt til å måle, så det kan slå negativt ut for egen videre karriere.

Disse forskerne uttrykker en form for ansvar og beskyttelse overfor dataene de produserer, og kan assosiere åpen tilgang med tap av kontroll over dataene. Dersom de ikke har rukket å kvalitetssikre dataene, kan forskerne bekymre seg for hvordan de blir gjenbrukt, og om de brukes innenfor de rammene og forutsetningene som ligger der.

5.3.5. Krav om datadeling fra tidsskrifter

Flere av forskerne innen fysisk oseanografi, samt forskeren innen paleoseanografi, nevner at det har blitt vanlig at tidsskrifter krever at data gjøres tilgjengelig sammen med publikasjonen av en artikkel. N9 forteller at «nå er det jo veldig ofte sånn at publikasjonen har dataene med og at det legges et sted hvor dette refereres til i selve publikasjonen.» Videre forteller forskeren at de som publiserer artiklene gjerne «har et krav om at hvis du lager en artikkel som er basert på noen data, så må også dataene følge med. Så det er en vanlig praksis nå. Da kommer jo publikasjonen og dataene ut samtidig.»

N9: Så er det jo også i mange tilfeller, som vi snakket om eksperimentelle og nye data som krever mer forskning før du vil legge dem ut. Der kan det gjerne ta flere år før forskerne synes at dette her er gode nok data, før de kan legges ut. Men her har det blitt krav om at du skal legge ut alle data med en gang, og da er du plutselig i en annen situasjon, for da vil du si: Nei, dette blir for risikabelt. Vi kan ikke bare legge ut nye data, uten å dokumentere de eller vite hvor gode de er, og så videre.

N9 sier at en konsekvens kan være at forskerne ikke vil bruke tidsskriftet som krever at man gjør dataene tilgjengelig sammen med publikasjonen.

N9: Hvis forskerne virkelig ønsker å publisere dataene sine, så gjør du det i et tidsskrift som kombinerer data og vitenskapelige resultater, det er samme tingen. Men hvis du

ikke ønsker å publisere dataene så vil du jo ikke bruke den journalen, da vil du gå til en annen journal.

Walters (2020) skriver i en artikkel at mange anerkjente tidsskrifter har etablert policyer som krever at forfatterne av artiklene gjør dataene som er brukt i analysene tilgjengelig for andre. Han mener det er to problemer med disse mandatene. For det første er det mange forskere som ikke klarer å etterleve og følge opp disse datadelingspolicyene, så selv om det er et krav i teorien, så fungerer det ikke nødvendigvis i praksis eller i virkeligheten. For det andre mener Walters at problemet med datadelingspolicyene for tidsskrifter, er at de ikke har noe virkelig insentiv for etterlevelse, siden belønningen for å publisere i et tidsskrift som krever datadeling, ikke er noe større enn belønningen for å publisere i et tidsskrift som ikke krever noen slik datadeling. Gitt den ekstra tiden og arbeidsinnsatsen som kreves for å gjøre dataene tilgjengelige, kan datadelingspolicyer for tidsskrifter, til og med oppmuntre til alternative strategier som ikke nødvendigvis er til det beste for den vitenskapelige kommunikasjonen. For eksempel kan krav til datadeling fra individuelle tidsskrifter oppmuntre forfattere til å sende arbeidet sitt til andre tidsskrifter som ikke krever datadeling, siden datadeling innebærer en kostnad uten noen tilsvarende fordel. Forfattere kan også vurdere å utsette innsendingen av en artikkel for å utnytte dataene mer fullstendig, før de deles med andre, og på den måten forsinke rapporteringen av potensielt viktige funn (Walters, 2020).

5.3.6. Hvilke data deles ikke?

Informantene innenfor samfunnsvitenskap har flere utfordringer knyttet til å dele data, enn naturviterne. Dette handler om at forskningen involverer mennesket som studieobjekt og egenskapene ved kvalitative data. Informantene innenfor samfunnsvitenskap viser til etiske årsaker til at ikke intervjudata kan deles utover egen forskningsgruppe. Det er vanligvis et krav at man anonymiserer gjenkjennelige detaljer, og at man behandler sensitive opplysninger med varsomhet. I intervjuundersøkelser som omhandler sensitive temaer, er det nødvendig å innhente samtykke fra informantene som skal intervjues.

S10: Vi som har med mennesker å gjøre – det er veldig strenge etiske retningslinjer. Så vi kan ikke dele på samme vis som naturvitere [...] det er viktig å beholde anonymiteten. [...] Og når man jobber med urfolk er det ekstremt viktig, og der har man en helt annen

type beskyttelse også, fordi det er så mange forskere som bare har kommet inn og tatt med ut. Derfor er det så viktig at data, - data kan ikke deles, men det kan presenteres tilbake.

Data som kan knyttes til enkeltpersoner, vil komme inn under virkeområdet til personopplysningsloven, og som hovedregel er det ikke rom for å bruke personopplysninger til andre formål enn det som er beskrevet i samtykket. Men med eksplisitt samtykke om deling, vil det være mulig å dele etter de vilkårene som er beskrevet. Men forsker S3 anser at dette kan forspille mulige kontakter og gjøre at folk ikke ønsker å delta.

S3: Jeg tror det er veldig mange intervjuobjekter som ikke ville akseptert et intervju, hvis det på en måte ligger til grunn at intervjuet i sin helhet er tilgjengelig for de som er interessert. Fordi det er jo akkurat den fortolkningen, eller den konteksten det finnes i som trengs, og forskeren har jo et ansvar der også, hvis hvem som helst kan lese det intervjuet i sin helhet 20 år senere.

S4 nevner også at dette «med kilder og kontakter er ganske følsomt».

S4: Det er klart at når det gjelder kilder, så er idealet at man oppgir, den konkrete referansen, ikke sant, hvis man har hatt et intervju. Det er det ideelle. Og da kan du si at dette er verifisert - og at en annen forsker har mulighet til å oppsøke samme person, og kan sjekke om dette faktisk er sånn. Det er det beste, men i virkelighetens verden, så er det ikke fullt så enkelt, en del intervjuobjekter ønsker ikke å bli sitert.

Forsker N9 påpeker at det er et EU-direktiv, at data skal være så åpne som mulig og at Norge er med og følger dette direktivet, men så «har du jo da spørsmålet om hvilke data som skal være åpne, og hvilke skal ikke være åpne, og hvilke delvis åpne under visse betingelser, så du har et spektrum av betingelser som gjelder for ulike typer data.»

N9: Hvis du bare hadde helt enkle klimadata som åpenbart skal være fritt tilgjengelig så er det uproblematisk, fordi at alle er enige om at disse dataene skal være åpne, og det er viktig at alle får tilgang til dem. Men så beveger du deg inn mot data som kan ha kommersiell interesse, som ressursutvinning, fiskeri og havressurser. Da er du med en gang over på data som ikke nødvendigvis skal være åpne for alle. Det har å gjøre med ressurser og økonomiske interesser og slike ting. Derfor begynner det med en gang å komme betingelser for hvem som kan bruke dem og sånn. Noen deler av dataene kan være åpne og så er det noen deler av dataene som ikke skal være åpne, og eller det er bare noen som kan bruke dem. Så har du også en annen økende problemstilling – dette

med sikkerhet og sikkerhetsaspektet, du har alt i fra trafikk og annen aktivitet som har med sikkerhetssystemet å gjøre, altså marinskipstrafikksikkerhet, så da har man en rekke betingelser på data som knyttes til sikkerhet på havet, sikkerhet for skipstrafikk og så videre.

En annen forsker viser også til at det finnes klassifiserte data som ikke skal være åpne.

N5: Det er en del klassifiserte data i batymetri. Hvis du måler havdybder i fjorder med en høy nok oppløsning, så er det klassifisert informasjon. Det er fordi at den norske marinen, tenker at den informasjonen kan noen bruke til å kjøre ubåter, og det er klart det vil man jo hindre dem i å gjøre.

Det finnes flere dilemmaer knyttet til å gjøre forskningen mer åpen. I en del sammenhenger kommer for eksempel hensyn relatert til sikkerhet, personvern og kommersialisering i konflikt med målsetningen om økt åpenhet. Dette kan føre til utfordringer med hensyn til forvaltning av data, og politikken på området har gått igjennom endringer. I starten var det viktig å få med alle miljøer på idéen om datadeling, og man innførte begrepet «åpne data» og konseptet «open by default». Dette ble møtt med en del kritikk, og konseptet har derfor blitt moderert. Nå er derfor budskapet at data og andre resultater fra forskning skal være «så åpne som mulig og så lukket som nødvendig» (Forskningsrådet, 2021, s. 13).

5.3.7. Dataenes «mobilitet»

Foruten om data som ikke deles av etiske, juridiske eller kommersielle årsaker, er det også enkelte typer data forskerne vurderte som lite egnet for deling. Holdninger til å dele data kan påvirkes av synet eller oppfatninger om hvor «delbare» data er, eller hvor gjenbrukbare disse dataene vil være for andre. Mens noen data kan fungere som «boundary objects», det vil si meningsfulle objekter som kan bevege seg mellom ulike brukere og forskningsprosjekter, samtidig som de beholder tilstrekkelig beviskraft (Star & Griesemer, 1989, s. 408), kan andre typer data, kreve ulike typer av faglig ekspertise, samt ulikt nivå av kontekstinformasjon – som for eksempel kjennskap til hvordan dataene ble skapt. Der forskerens formål med å gjenbruke dataene også kan spille inn på hvilket nivå – eller hvor inngående dataene må forstås.

Mer kompliserte data vil kreve mer kontekstinformasjon, enn enklere og mer standardiserte data for å være gjenbrukbare. Vanskeligheter med å trekke ut data fra selve forskningskonteksten er kjent som «the mobility problem». Å gjøre data mobile slik at de skal

kunne gjenbrukes av andre enn dem som frembragte dem, krever at man skiller dataene fra den forskningssituasjonen der dataene ble til som data. Som et resultat vil noe av meningen gå tapt. Å dokumentere dataenes proveniens kan gjenopprette noe av denne meningen, men den presise konteksten vil aldri være fullt ut overførbart (Borgman, 2015, s. 219).

Forskningsrådet påpeker at forskningsdata som regel må utstyres med rike metadata for at dataene skal gi mening og kunne gjenbrukes av andre. Dette gjelder spesielt forskningsdata som ikke er del av store datainnsamlinger hvor infrastruktur og etablerte standarder og formater er på plass (Forskningsrådet, 2021, s. 4).

N11 forteller at forskere kan ha brukt årevis på å fremskaffe data, og som de da kjenner svakhetene og styrkene til. Derfor foretrekker vedkommende å samarbeide med de forskerne som har fremskaffet dataene «de kjenner dataene godt, og noe er nedfelt i notater, men så husker du også at den tirsdagen der, så var det faktisk overskyet, eller den dagen der gikk noe galt, så de har mye forklaringer rundt. Så det er ofte mer ved dataene, enn dataene».

Denne forskeren uttrykker en viss skepsis til å gjenbruke data, som man selv ikke har vært med på å fremskaffe, fordi man ikke kjenner dataene godt nok. Borgman påpeker at data eksisterer i en kontekst og at «data are not pure or natural objects with an essence of their own (Borgman, 2015, s. 18). Videre at de eksisterer i en kontekst, og tar opp i seg mening fra denne konteksten og fra betrakterens perspektiv. I hvilken grad disse kontekstene og betydningene kan representeres, har innflytelse på overførbareheten av dataene. Det er ikke før kontekstualisering skjer under forskningsprosessen at data får sin egen mening (Borgman, 2015). Betydningen konteksten har for forskningsdata har vært påvist i tidligere studier (Borgman 2015, Pryor 2009) og diskusjonen rundt dataenes kontekstualitet kan summeres opp med to slags typer kontekst (Kurata et al, 2017, s. 2). En er den «eksplisitte» konteksten, slik som for eksempel metadata, mens den andre er den «implisitte» konteksten, som også kan inkludere taus kunnskap (Kowalczyk & Shankar, 2013). Den beskrivende «eksplisitte» konteksten refererer vanligvis til metadata som kan legges til for å beskrive forskningsdataene og som er nødvendige for at dataene skal kunne gjenbrukes (Kurata et al., 2017, s. 2).

I tillegg baserer forskningspraksiser eller aktiviteter seg ofte på uformell eller taus kunnskap, som ikke alltid er enkel å dokumentere, og som i stor grad baserer seg på implisitt kunnskap delt på varierende nivåer av forskningsfellesskap. Hvorvidt forskere kan tolke dataene som produseres av andre, avhenger av tilgjengeligheten til disse implisitte kontekstene. Selv om den «enkle» handlingen med å dele data, ofte er ment å være nok til å gjenbruke dataene, har forskere gjentatte ganger uttrykt bekymring for at mangel på kontekst ville hindre dem i å dele og gjenbruke data enten som produsenter eller gjenbrukere. Noe som kan resultere i at dataene kan mistolkes eller brukes utenfor sine forutsetninger (Faniel & Jacobsen, 2010).

N11: Hvis du har funnet ut noe robust, så skal det ikke endre seg ved å analysere det på nytt, men det kan jo være at ikke alt av støttematerialet er tilgjengelig, sånn at det er vanskelig å lese datafilene. Og det er jo nettopp det vi jobber med da, sånne datamanagement plan, det er å sørge for at de metadataene, de som forklarer hva slags data dette er og hva er rammene for bruken, slik at det er tydelig og at det ikke brukes utenfor, men selv om vi har med det, så tror jeg ikke det er nok egentlig. [...] Litt tilbake med dette med kvalitetssikring, så synes jeg egentlig det er lurt å kontakte eierne av dataene, som har litt historikken og sånne ting.

Når forskere gjør data tilgjengelig for andre, kan ikke en forsker alltid kommunisere helheten i konteksten i hvordan dataene ble produsert, og det kan være vanskelig å tolke og gjenbruke dataene korrekt uten denne implisitte informasjonen. Selv om datasettene som blir delt i en database blir utstyrt med metadata, gir noen forskere uttrykk for at dette ikke alltid er nok.

N8: Altså hvis du laster ned data, så er det ikke alltid at det er like lett, selv med metadataene, og vite hva som er gjort fra dataene er tatt, til sluttproduktet, og eventuelle usikkerheter i dataene. [...] Av og til så kan det jo være sånn at hvis du laster ned et datasett fra nettet, en annen partner, så tenker du - jo disse dataene er 100% ok, men hvis du jobber med data selv, så vet du litt mer om usikkerheten.

For å kunne gjenbruke dataene korrekt, og innenfor de forutsetningene som ligger til grunn for dem, så vil det ofte kreve at man har innsikt i de prosessene som går forut for opprettelsen eller dannelsen av dem (Pasquetto et al., 2019).

N1: Jeg har gjort eksperimenter, og de dataene er veldig vanskelige for noen som ikke har vært med på eksperimentet, den er helt ubrukbar, for de som ikke har vært med på eksperimentet.

Selv om dokumentasjon som metadata legges ved, kan det ofte være slik at mye av den kunnskapen som trengs for å forstå et datasett er taus. Det vil ikke alltid være mulig å uttrykke denne kunnskapen eksplisitt, slik at andre skal kunne forstå og ta i bruk dataene for gjenbruk (Birnholtz & Bietz, 2003).

5.3.8. Dataskaperens fordel

I noen tilfeller kommer kontekstinformasjonen forskerne trenger fra metadata, i andre tilfeller blir denne kontekstinformasjonen delt og utvekslet gjennom å ta del i forskerfellesskap (Faniel & Jacobsen, 2010). Enkelte ganger snakker forskerne direkte med kolleger for å forstå hvordan dataene er produsert (Birnholtz & Bietz, 2003).

N9: Vi opplever av og til at vi får henvendelser fra andre som ønsker å vite mer om dataene, og da er det jo dette med at man har god dokumentasjon på dataene. Det er viktig, slik at de som skal bruke datasettene - kan finne ut det de trenger å vite om det datasettet. Du har gjerne referanser til publikasjoner om datasettet, slik at det skal være så åpent og tilgjengelig som mulig. Selvfølgelig, noen ganger er ikke dataene godt nok dokumentert, så det er ikke lett for en utenforstående å bruke de dataene, uten at de må ha betydelig hjelp.

Pasquetto et al., viser også til at dataskaperne av et datasett kan ha en fordel, fordi de kan ha personlig og taus kunnskap, som kan brukes som byttehandel for å danne samarbeid til gjensidig fordel for både den som har skapt datasettet og den som ønsker å gjenbruke det (Pasquetto et al., 2019).

N9: Hvis det er kompliserte data må de ha hjelp av forskerne for å kunne bruke dataene videre. Så det er en typisk ting ved kompliserte data, at vedkommende som skal bruke dem, vil trenge hjelp til å bruke dataene.

Potensielle datagjenbrukere fyller ofte hull i kunnskapen sin ved å be om hjelp fra dataskaperne for å kunne gjenbruke dataene, til gjengjeld kan de som gjenbraker dataene ofte kreditere dem som medforfattere (Pasquetto, 2018, Wallis et al., 2013).

5.3.9. Forskningsdata – hva slags produkt skal man dele?

Betegnelsen rådata brukes om data i sin opprinnelige form, slik de er når de samles inn. Dersom slike data systematiseres, sorteres eller analyseres, går de over til bearbeidede data (Forskningsrådet, 2021, s. 12). Forskningsdata kan gjennomgå ulike stadier av transformasjon fra de opprinnelige rådataene, og data kan bli gjort tilgjengelig for andre forskere på ethvert stadium i prosessen. I forskningsmiljøer kan det også være diskusjoner om hvor bearbeidede dataene som skal deles bør være. En forsker viser til at det kan være risikoer forbundet med å legge ut rådata.

N8: Mens data som blir samlet inn fra båt og sånn, da har du vel utfordringen med at det du samler inn på båten, altså rådataene, det er ikke meningsfylt å bare legge ut alle data. Det blir bare masse rot, og andre folk som skal bruke dataene har ikke forutsetningene til å få noe fornuftig ut av dem, og da risikerer du at de blir brukt feil. Så da er utfordringen å legge ut nok data, men at du har bearbeidet dem først.

N8 forteller at det kan oppstå diskusjoner i noen prosjekt «på hva slags nivå skal disse dataene gjøres tilgjengelig?» Og at det kan være en utfordring, der ulike partnere er involvert «ikke bare et universitet, men flere universitet og ulike partnere, og ulike instruksjoner og ulike måter de pleier å bearbeide data på. Så da er det jo en utfordring for alle å bli enige om hva slags produkt som skal legges ut, og når tidsmessig.»

N8: Jeg tror risikoen er at hvis du publiserer alle dataene, på en eller annen måte som gjør at det ikke er helt klart for [...] hvordan de dataene skal behandles og bearbeides, og at du da kan risikere at de som ikke har kjennskap til disse dataene kan tolke dataene feil rett og slett, og at du da kan få resultater som er feil. Ikke på grunn av at disse forskerne gjør noe feil med vilje, men rett og slett, det er ikke alle data som er like lette og bare laste ned. Hvis du laster ned temperatur for Oslo, et punkt for hvert år, så er det ganske lett å tolke, men hvis du har mer kompliserte datasett som ikke er bearbeidet, så tror jeg du kan, i teorien skape vanskeligheter for de som skal analysere dem.

Men dette kan forholde seg ulikt for ulike typer data. Det var andre informanter som viste til at det gjerne var rådataene de ville ha tilgang til eller ble spurt om å dele. N6 forteller at det som oftest er rådata man blir spurt om å dele, «så det er oftest mer rådata, enn bearbejdede data.»

N2 forteller at «utfordringen kan være [...] at de publiserer en tolkning, men har glemt å publisere selve rådataene, som går forut for denne tolkningen». Og at det kan være viktig å få tak i selve rådataene, for «da kan man gjøre beregninger på en annen måte. [...] Da fins det masse andre – jeg skulle kunne kjørt igjennom det med andre program.»

5.3.10. Datadeling på forespørsel

Ved siden av den mer formelle formen for datadeling, som datadeling ved publisering av artikler og datadeling via databaser eller datasentre, så er en annen vanlig form for datadeling – datadeling på forespørsel. Dette er en uformell form for datadeling (Goodwin, 2012). Ved gjennomgang av intervjuene ser det ut til at denne formen for datadeling er den vanligste. Selv for de forskerne som deler mye data via databaser eller datasentre, virker denne formen for datadeling å være minst like vanlig.

N7 forteller at «som regel spør man andre forskere» om data, og at vedkommende ikke selv har lastet ned noe fra arkiv. Ofte så er det «fysiske tidsseriedata», og det er gjerne fra «folk man har en viss kjennskap til».

N1 opplever at folk tar «kontakt på et tidligere stadium, innen data ligger i arkiv».

N9 forteller at «det er jo ofte sånn det har vært tradisjonelt at en forsker har noen data, så annonserer han at han har disse dataene, og hvis noen er interessert i å bruke dem så bare henvender du deg. Så vet forskerne hvem som har brukt dataene, for det er liksom en del av selve forskningsprosessen.»

Dermed kan forskerne som deler dataene, vite hvem som har brukt dataene og hvordan de har blitt brukt av andre. Blant noen forskere kan dette være den foretrukne måten å dele data på, for å kunne sikre at dataene blir brukt på riktig måte. Datadeling på forespørsel har den fordel at forskerne kan ha direkte kommunikasjon seg imellom. Thoegersen & Borlund (2022), viser til at forskere kan ha ulike holdninger til ulike former for datadeling. Som nevnt

tidligere ville forsker N11, dele data eller gjenbruke data som en del av et samarbeid med de opprinnelige eller opphavelige dataskaperne, blant annet for å sikre at dataene kan bli forstått og blir brukt på rett måte.

Men selv der deling av datasett har blitt gjort tilgjengelig i databaser eller arkiv, blir forskerne ofte kontaktet personlig. Blant annet fordi de som ønsker å gjenbruke dataene, ikke selv har funnet frem til dem.

N8: Men problemet er at noen ganger blir disse dataene gjemt inne i en databaseverden, som ikke er veldig intuitiv, som gjør at når folk sender e-post til meg og spør om de kan få de dataene bla, bla. Hadde de dataene vært lettere tilgjengelig: Jeg kan sende en link her er de, men de har ikke funnet dataene selv, fordi de er ikke synlige nok.

N2: Det hender at folk tar kontakt og ber om data, som faktisk ligger ute i arkiv, men det vet ikke de, fordi de har ikke funnet dem. Det er veldig vanlig at man skriver til andre forskere og spør om man kan få dette datasettet, men har kanskje prøvd å finne det, av og til kan det være vanskelig å finne ting.

Dette viser at datasett som ligger i dataarkiv eller databaser, kan være vanskelig å finne frem til.

5.3.11. FAIR-data

For at dataene skal være gjenbrukbare må de som nevnt tidligere i teoridelen, være tilstrekkelig dokumentert og det må registreres så fullstendige metadata som mulig. I *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata* (2017), påpekes det at Norge vil slutte seg til EUs oppfordring om å følge FAIR-prinsippene. Strategien påpeker også viktigheten av langsiktig arkivering og kuratering av forskningsdata, og at dette arbeidet må være basert på internasjonale standarder. FAIR dreier seg i hovedsak om å håndtere og beskrive data på en måte som setter andre i stand til å finne, forstå og bruke dem.

N9 forteller om utfordringer med kravet om å gjøre dataene FAIR:

N9: (...) standardisering og etablering av metadatasystemer er jo et stort spørsmål og der utvikler det seg gradvis, men mengden data vokser gjerne fortere enn man er i stand til å systematisere dem, og legge dem inn i standardiserte arkiver. Så det er en stor utfordring. Det er kanskje den største utfordringen, det å organisere dataene på en sånn måte at det er lett å bruke dem i ettertid, og at man kan finne frem.

Et viktig aspekt ved datadeling dreier seg om hvorvidt data kan bli funnet, vurdert og forstått, og gjenbrukt av andre enn den eller de som opprinnelig frembragte dataene. Men for at data skal være FAIR, må de ikke bare være gjenfinnbare og gjenbrukbare for mennesker, i tillegg må både data og metadata kunne håndteres maskinelt.

N9: Nå er det jo snakk om dette med maskin til maskin, maskiner skal kunne finne data for deg, dette som kalles interoperabilitet mellom ulike datasystemer, det er det et krav til, men det fungerer bare i den grad man har standardiserte formater. [...] så kreves det ressurser for at data skal bli lagt inn i disse systemene, og at det gjøres tilgjengelig på en brukervennlig måte. Tradisjonelt så har jo datasentrene gjerne vært slik at de ikke er så brukervennlige. Datasentrene har vært laget av de som er eksperter på datasentre, de har lagt til noe som de mener var en god løsning for ti år siden eller sånn, men i dag stiller man nye krav som gjør at datasentrene ikke er så brukbare lenger. Så man må tenke litt nytt, man må tenke at datamaskiner skal snakke med hverandre og sånne ting. Det krever at man lager dette på en ny måte, og det skjer også rundt omkring, men her går utviklingen veldig fort, og forskerne - de som driver med sin faglige forskning, dette er jo ikke deres felt.

Det å gjøre dataene FAIR tar gjerne mye tid, og kan være ressurskrevende for forskerne. I tillegg viser N9 til at det er et behov for å utvikle de teknologiske systemene, og at de ikke nødvendigvis er så brukervennlige. Nye krav til deling av data og at data som deles skal følge FAIR-prinsippene, gjør at teknologien må utvikles, men her henger teknologien etter. Dersom forskerne skal kunne følge opp kravene om å gjøre dataene FAIR, må også teknologien utvikles. Dette er et felt som krever egne dataspesialister, som kan samarbeide med forskerne som skal tilgjengeliggjøre datasettene, slik at dataene kan bli tilgjengeliggjort så de kan møte kravene om å være FAIR. Samtidig bør de vite noe om behovene som potensielle gjenbrukere har for at de skal kunne anvende dataene. Forskeren viser også til at det kan være en interessekonflikt mellom ulike grupper eller aktører.

N9: Dette blir et felt som krever spesialister, og forskeren er interessert i å gjøre sine ting, så kommer de som skal drive med datasystemene, de som også skal bruke dataene, men det blir to forskjellige grupperinger, så det er et spørsmål om hvem som gjør hva,

og hvordan møtes de. Da har det gjerne en tendens til å bli et gap mellom disse gruppene.

Forskningsdata kan ofte være komplekse, og det å gjøre dataene anvendelige for andre, krever både tid og kompetanse. Denne kompetansen har ikke nødvendigvis forskeren som har produsert dataene. Forsker N7 viser til at i tillegg til gode teknologiske løsninger, så mangler det også menneskelige ressurser som kan bistå forskerne med å gjøre data tilgjengelige.

N7: Igjen det går på det praktiske. Jeg synes at i praksis, så er vi ganske dårlig skodd for å i realiteten gjøre data tilgjengelig, og gjøre de brukbare og FAIR. Så det er hele tiden de praktiske hindrene, jeg har ingen moralske, etiske eller faglige begrensinger, eller grunnlag for å ikke dele, men det praktiske. Det går på alt fra tilgang til lønnsmidler, forskningsteknikere som datamanagere, som faktisk kan gjøre disse analysert og tilgjengelig, til de mer basale behovene som tilgjengelige datalagringssystemer. Vi har veldig dårlige tekniske nasjonale løsninger. [...] Det tar mye tid og ressurser å gjøre dataene ... og standardisere dem, og vi har ingen nasjonale gode arkiver, som faktisk gjør dataene ordentlig tilgjengelig. Du skal være ganske dyktig innen fagfeltet for å kunne identifisere dataene og bruke dem.

N7 mener det må sørges «for at vi har nasjonale infrastrukturløsninger som gjør dataene reelt tilgjengelige».

Det å gjøre data FAIR fordrer ikke bare ressurser i form av investeringer i gode infrastrukturer og systemer for deling og gjenbruk, det kreves også investeringer i form av menneskelige ressurser. Det er derfor et behov for kompetent forskerstøtte, som kan veilede og gi råd til forskere, for å nå målet om mer deling og gjenbruk av forskningsdata.

N7 forteller at forskningsdata i dataarkiv ofte er «det veldig komplisert å finne dem igjen, og det er software som er fryktelig utdatert».

N7: Det er fravær av standardiseringer og ikke minst kanskje, og det viktigste er at det er systemer som ikke lar seg kombinere med maskinlesing [...], så du må grave deg ned i de enkelte arkiver. Fordi data ikke tillater, ikke er kompatible med operativsystemer og frontend-løsninger som brukes.

5.3.12. Mangel på tid og ressurser

Som jeg har vært inne på tidligere i denne oppgaven, så er tid en faktor som kan virke inn på deling av forskningsdata. Deling er også et ressurs spørsmål, man kan ha finansiering til det man selv skal gjøre, men ikke til å legge til rette for andre. Å dokumentere data kan bli sett på som en byrde, delvis fordi å dokumentere data for offentlig eller andre forskeres bruk, er ulik den dokumentasjonen som trengs for personlig bruk (Faniel, 2009).

N9: Jeg tror forskerne er interessert i at dataene deres skal bli brukt av andre, men det er igjen den der jobben med å gjøre det tilgjengelig. Det er en flaskehals i det systemet som vi har nå. Det er at det er noen som produserer data på den ene siden og de skal da tilrettelegge for at andre skal bruke dem, men den jobben imellom de to leddene - hvem skal gjøre den jobben? Og vanligvis har ikke forskningsprosjektene satt av noen midler til at man skal gjøre den jobben i ettertid. Derfor blir det ofte et gap der, og det er en del data som blir liggende hos forskerne i istedenfor at de legges inn i disse datasystemene.

Fordelen av en slik innsats går i stor grad til sekundærbrukeren. Flere av forskerne peker på at det ikke er satt av midler til å gjøre den jobben som kreves for å tilgjengeliggjøre dataene for andre.

N7: det å stille til rådighet lønnsmidler, for å gjøre dataene til de som ønsker og de som er villige til å gjøre data tilgjengelige, burde det vært bedre nasjonale ordninger, for å faktisk hjelpe til at de deles.

Det kan være store kostnader knyttet til å gjøre datasett tilgjengelige, fordi det både er tidkrevende og arbeidsomt, å tildele metadata og annen dokumentasjon, slik at data skal kunne gjenbrukes av andre.

N6: min utfordring er jo mere det at, det er ikke det at jeg ikke vil dele data, men det kan være så mye å gjøre at jeg ikke får publisert dem. At man ikke har nok tid til å publisere data.

N6 forteller blant annet at «det jeg er avhengig av, er at det er datafolk som sitter - som kan faget og som kjenner hva jeg trenger og kan bidra til at jeg får dokumentert dataene på en god nok måte, til at jeg kan få publisert dem».

Flere av forskerne etterlyser mer veiledning og assistanse for å gjøre dataene sine tilgjengelig.

N8: Jeg tror ikke hindringen er at folk ikke har lyst [...]. Forskere generelt har veldig mye å gjøre, og det man har lyst til å gjøre er den forskningsbiten, og da blir datalagring, sånn teknisk ekstra jobb, som du føler blir merarbeid, og derfor blir det skjøvet til siden. Og så må du vente eller du må, eller du er litt avhengig av at noen jobber spesifikt med sånne ting, og nesten banker på døra di og sier nå skal vi gjøre dette her, nå skal jeg vise deg, nå skal jeg hjelpe deg. Så jeg tror av og til så er det rett og slett bare arbeidet.

Å dokumentere data tar mye tid og krefter, og opplæringen og verktøyene som kreves for å støtte en slik innsats er ofte mangelvare (Baker and Yarmey, 2009).

N9: Man kan gå ned på et første nivå, det er at et prosjekt produserer data, så lagrer prosjektene de dataene de har på en server, gjerne i formater som er etablert innenfor feltet. Men så ønsker man gjerne at de dataene, når prosjektet er ferdig inn i et nasjonalt eller internasjonalt datasenter, så det er en vanlig måte å gjøre det på. Men da må man ha, det er gjerne arbeid knyttet til å gjøre den jobben og hvis prosjektet ditt er avsluttet så er det spørsmål om hvor mye tid skal du bruke på å tilrettelegge data for at andre skal bruke dem.

5.3.13. Viktigheten av anerkjennelse og kreditering, og mangel på insentiver

Tidligere forskning har vist til at forskere ikke har en innarbeidet kultur for å sitere datasett. Standarder for datasitering er blitt utviklet, men praksisen med å sitere datasett er fortsatt ikke helt på plass (Wallis et al., 2013).

Mangel på anerkjennelse som kreditering og sitering, kan virke som et disinsentiv for å dele forskningsdata. Dersom den som deler opplever at det ikke blir tilstrekkelig verdsatt, vil det være demotiverende å dele data.

N9: Nei, det er jo bare det at forskere som har opplevd at dataene har blitt brukt uten å bli referert til så vil de jo være veldig forsiktig med å gjøre dataene åpne [...] man må vite at hvis man gjør dataene tilgjengelige, så blir de riktig referert av dem som skal bruke dem. Hvis det ikke skjer, så er det ikke fristende for forskerne å legge ut dataene sine. Da vil de heller si «nei, jeg vil bruke dataene selv, og dele dem med kollegaer som jeg stoler på, men hvis de skal legges åpent ut, så vil jeg være litt bekymret for at de kan bli misbrukt. [...] Akkurat som annen forskningspraksis – du refererer til andre papere når du skriver om noe, du henviser til hva andre har gjort, så skal du alltid ha en referanse til hvem sitt paper du refererer til. Det er akkurat det samme med data også, du må referere til de personene som har produsert dataene, hvis du skal bruke de dataene.

N9 forteller at «hvis forskerne som egentlig har levert dataene, ser at her er det noen andre som setter sitt navn på det, uten å referere til dem, så vil de si at de har misbrukt de dataene.»

N9: Vi har sett i mange år at det er veldig lett å hente inn data fra en database og bruke dem videre, uten å referere til hvor dataene kommer fra eller hvem som har produsert dataene.

Siteringer nevnes som en måte å få anerkjennelse for det tidkrevende arbeidet, det ofte er å produsere data, samt gjøre disse tilgjengelig for andre. Å få anerkjennelse for dette arbeidet oppleves som viktig.

N8 forteller at det som kanskje har vært problemet før, er at du kan jo bruke mye tid på et datasett, men så får det ingen synlig belønning. Men nå kan datasett bli publisert med en DOI, og det kan i alle fall kan telle som et arbeid. «Så det tror jeg, at det at du kan sitere det publiserte datasettet, det vil jo hjelpe.»

Forskeren H12, forteller at vedkommende har opplevd at datasettene har blitt sitert, men at det fortsatt ikke er så vanlig. Et pluss er at «da får man få sett hva ens data også kan brukes til, i tillegg til det man selv bruker det til [...] og da tenker man videre, og man kan komme med nye spørsmål til sine data, og dermed kan man komme frem til nye tolkninger».

Forsker N6 mener at det å publisere data, spesielt med DOI, bør få høyere status og kredibilitet.

N6: Jeg tenker at det å dele datasett, og spesielt med en DOI og ikke bare legge ut en excel-fil, men publisere det som ordentlig data må få høyere status. Så det at når du setter på cv'en din hvor mange artikler du har publisert, så skal du også sette på hvor mange datasett du har publisert. Og at det får litt kredibilitet. Jeg tror det er viktig for de som jobber med å telle tellekanter, at det blir brukt og at det blir referert til og at det blir større forståelse for at det å produsere og tilgjengeliggjøre datasett på en god måte er noe som er viktig for forskningsmiljøet, og at det teller på en måte.

N5 forteller imidlertid at:

N5: Man skulle jo egentlig kunne sitere datasett, men det blir i praksis ikke gjort hvis det finnes en artikkel som beskriver de dataene. Så vil man jo ... jeg tror det skal mye til at det kommer til å endre seg. En publikasjon er jo kvalitetssjekket med fagfelleevaluering og et datasett er jo bare en haug med målinger. De kan være feil.

En annen barriere er at det ikke finnes noen egentlige insentiver for å dele data.

Publiseringsindikatoren som er en del av grunnlaget for tildeling av midler til forskningsinstitusjoner, gir hovedsakelig uttelling for publiserte artikler, og ikke dataene som ligger til grunn for dem.

N1: «Ja, i det her tellekantsystemet, som vi liksom lever i. Du rapporterer fint inn dine artikler, men det å publisere datasett, det bryr de seg ikke om. Som forsker blir du vurdert på dine publikasjoner, og H-index, men å publisere datasett syns ingen steder. Jeg publiserer mine data uansett, for jeg innser hvor viktig det er at man gjør det, men det er jo litt surt å ikke få mer kred for det, for det er mye jobb. Mye.»

N9: «Forskerne blir ikke belønnet for å gjøre dataene tilgjengelig. Du blir belønnet for å lage en publikasjon, og nå hvis publikasjonen krever at dataene gjøres tilgjengelig, så må du jo gjøre det, men de fleste publikasjoner er ikke slik, der er det bare de vitenskapelige resultatene som teller og dataene de ligger i bakgrunnen, men verdiene av de dataene blir ikke tilstrekkelig – skal vi si vurdert, så det er et gap der da. Men vi ser at utviklingen går mot at dataene må - altså man må belønne de som produserer dataene, og de må bli referert og så videre sånn.»

Forskerne opplever en missing link – at de ikke blir belønnet for å gjøre dataene sine tilgjengelige. De blir belønnet for å lage en publikasjon, og hvis forskningsråd eller tidsskrifter krever at de gjør dataene tilgjengelig så må de gjøre det, men det er bare de vitenskapelige publikasjonene som teller. Verdien av de dataene som gjøres tilgjengelig blir ikke tilstrekkelig vurdert, selv om de som produserer dataene skal bli sitert, er det ennå ikke en helt innarbeidet praksis. Det eksisterer et gap mellom de som produserer data på den ene siden, og som da skal tilrettelegge for at andre skal bruke dem og de som skal bruke dataene på den andre siden.

5.3.14. Hvorfor deles data?

Forskere produserer store mengder data, der noen data kan være nyttige for andre. Til tross for alle barrierene med å dele data, inkludert betydelige kostnader – både i arbeidsinnsats og tid, blir deling av data oppmuntret fordi vitenskapelige og sosioøkonomiske fordeler, er antatt å være betydelige (Zimmerman, 2003). Dette kommer tydelig frem i de mekanismene som blir skapt for å støtte datadeling – som for eksempel infrastrukturer, og fra retningslinjer og

mandater som oppmuntrer til deling av forskningsdata. Alle disse aktivitetene er intendert for å oppnå visse ønskede resultater (Zimmerman, 2003).

Flere av informantene har vist til at tidsskrifter og finansieringskilder krever at datasett deles ved publisering, særlig gjelder dette innenfor geovitenskap og geofysikk. Men også N7 sier det har hatt betydning. Når jeg spør N7, om policyer fra forlag eller finansierer har hatt betydning for at du har delt data? Sier informanten: «Jo, det har det. Spesielt finansieringskilder, hvor det stilles strenge krav, så det har jo betydning. Det har gjort at man har speedet opp prosessen.»

Flere av informantene delte oppfatningen om at data som er finansiert med offentlig midler, bør deles. Flere deler grunnsynet om at data som er finansiert av det offentlige ikke er personlig eiendom, men et kollektivt gode, som flere skal kunne bruke.

H12: Jeg mener det er helt essensielt for å kunne bidra i forskningen og kunne komme videre i forskningen at man deler data. [...] Og det er litt sånn at hvis det er offentlig finansiert forskning, så kan man ikke tillate seg å drive med det for sin egen del. Da har man et samfunnsansvar som innebærer deling.

N5: Alle vi yngre - vi ser både at siteringer til de som deler data blir jo god, pluss at det er jo staten som har finansiert dette her. Så det skulle bare mangle at ikke ... Hvis det er et firma som betaler lønnen, så kan man jo forstå at man bør ..., så er det jo eiendommen til det firmaet, for jeg gjør den jobben for det firmaet. Men hvis det er staten så bør det komme alle til gode.

Flere av informantene fra naturvitenskap delte også synet om at data som kan knyttes til en publikasjon skal være tilgjengelig på forespørsel. Ifølge Zimmerman «The norms of science dictate that data associated with a publication should be available upon request» (Zimmerman, 2003, s. 194).

N6: Du må vente på publisering, men når det er publisert, så skal du ha det. Da er det aldri noe spørsmål. Da er det ikke noe problem å få tak i, eller å dele.

Selv om denne normen blir avvist en gang iblant, er det en gjenkjennelig etisk vitenskapelig norm.

N2: Kan du sende meg det her grunnleggende datasettet, fordi jeg vil teste andre måter å beregne datasettet på? Da fikk jeg til svar – «Nei, dette er noe vi fortsatt holder på med, og derfor kan vi ikke ...». Det er jo feil, det her er publiserte data, og jeg skal kunne be om dette.

En motivasjon for å dele data, kan være at det leder til flere publikasjoner.

N2: det kan jo være at noen andre ser mitt datasett, blir interessert og tar kontakt med meg, og at det kan lede til publikasjoner. [...] Så jeg tror at å dele data ofte gir mulighet for flere publikasjoner for ens egen del.

Flere av forskerne mente det var viktig å dele data for å kunne se ting i en større sammenheng.

N6: det er kjempeviktig, fordi man klarer ikke å se det hele, eller det store bildet, uten at man har mange typer små datasett som man setter sammen, du mister veldig mye hvis man ikke deler små datasett, da blir alt veldig stykkevis og delt, og veldig spesialisert og fokusert, så de større bildene som kan komme opp, når man for eksempel ser på data fra ulike perspektiv, eller ulike typer data som er satt sammen, du får en helt annen potensiell innsikt i det, enn om man bare sitter med sine egne små data.

En påpekte at jo flere måledata du har for et punkt eller et geografisk sted, jo mer nøyaktig blir totalbildet ditt. Derfor er det viktig å dele data.

N8: Du har dine data som du har brukt på din måte. Jeg kan ha mitt forskningsspørsmål og samle inn data, men andre folk kan jo se på mine data på nye måter, og kan sette det i en større sammenheng. Så har du jo et godt eksempel med værvarsling, jo flere data du har jo bedre værvarsling får du, får den værvarslingsmodellen trenger ... Det er derfor ikke alltid at værmeldingen er så god hvis du er i en rar liten dal i Norge, på grunn av at det er vanskelig å få data derifra. Så det kan du si gjelder for mange andre typer målinger også, f. eks hvis du måler havtemperatur, jo flere plasser du måler, jo bedre blir totalbildet ditt, da er det viktig å samarbeide, og gi hverandre tilgang.

Noen nevnte at det var viktig å kunne sammenligne sine egne data med andres, for eksempel for å sjekke om andre har kommet frem til lignende resultater.

N11: Jeg tenker i mitt felt som i alle andre felt er det kjempeviktig. [...] Og siden det dreier seg om forurensning så er det kjempeviktig å få sammenlignet det med andre lab'er, og få en kvalitetssjekk på at det man har gjort er riktig». [...] Sånn som nå er det en ny kjemikaligruppe som det jobbes veldig mye med, og som det er ekstremt vanskelig å trekke konklusjoner, fordi det er vanskelig å sammenligne våre studier, med andre sine studier. Og hvis du da ikke har data så gjør du det bare enda vanskeligere.

N5: Du har jo nesten alltid et behov for å sammenligne dine egne data med noen andres.

En forsker nevner i tillegg verifisering av funn.

N8: [...] både så folk kan sjekke de resultatene og funnene som ble gjort, men også ta de dataene og bruke de videre og bruke dem i andre sammenhenger. [...] Både verifisering av funn, det er jo en ting, Men nesten det viktigste er det der med at du får satt ting i en sammenheng.

En nevnte at man kan spare mye tid for andre.

S10: Det å publisere datasettene sine tror jeg er kjempeviktig, man kan spare mye tid for et annet menneske.

En forsker viste til at det å publisere data kan føre til at man får høyere siteringsindeks på tilhørende artikkelen.

N5: Vi er jo interessert i å gjøre de dataene tilgjengelige, fordi hvis folk bruker dataene dine, vil de sitere artikkelen, og da får du høyere siteringsindeks.

En forsker påpekte at det ikke alltid er lett å få tak i data fra andre forskere privat, og at det kan være mer effektivt å få tak i data via datasentre.

N8: Det er ikke sikkert den forskeren svarer på e-post, det er ikke sikkert forskeren jobber med de dataene lenger. Da har du den utfordringen – det å få ting lagt ut på nettet, på slike servere og sånn, det er en grunn til at det har blitt mye mer vanlig de siste fem årene. [...] Det kan være lettere å få tak i data fra NASA, enn fra min egen institusjon.

De viktigste årsakene til å gjøre forskingsdata tilgjengelig var å få satt ting inn i et større perspektiv og en sammenheng. Å kunne sammenligne sine egne resultater med andres, for å sjekke om det man selv har kommet frem til er sannsynlig. Det anses som god forskningspraksis å dele data med andre, spesielt dersom forskningen er finansiert med offentlige midler. I tillegg til at det er viktig for å komme videre i forskningen.

6. Oppsummering av funn og diskusjon

Problemstillingen for denne oppgaven er: Hvilke holdninger og erfaringer har forskere til deling av forskningsdata? I tillegg har jeg formulert følgende forskningsspørsmål: Hvilke faktorer kan ha innvirkning på forskeres datadelingspraksiser?

Min undersøkelse er kvalitativ og speiler bare de tolv informantene sine svar. Det er derfor ikke alle tenkelige svar som er representert, og det kan også tenkes andre faktorer eller forhold som ikke kommer frem i intervjuene med informantene, og som kan virke inn på deres datadelingspraksis.

I denne delen vil jeg gi en oppsummering av funn og resultater, som samtidig besvarer den aktuelle problemstillingen og forskningsspørsmålet. Deretter avrundes oppgaven med en diskusjon, der jeg ser nærmere på deling av forskningsdata ut ifra et sosioteknisk perspektiv. Til slutt vil jeg si noe om begrensninger ved denne studien og gi forslag til videre forskning.

6.1. Sammenligning og kobling mot problemstilling og forskningsspørsmål

Jeg har intervjuet tolv forskere der åtte forskere har naturvitenskapelig bakgrunn og fire forskere har bakgrunn fra samfunnsvitenskap/humaniora. Beslutningen om å intervju forskere fra ulike fagdisipliner, og med ulike typer data, baserte seg på et ønske om å få med ulike perspektiver på deling av forskningsdata.

Forskningsmetodene forskerne bruker og deres studieobjekt, former ikke bare forskernes datapraksiser, men de former også dataenes karakter, eller egenskaper ved dataene, som kan innvirke på forskernes datadelingspraksiser. Hesse et al., poengterer at «in all scientific disciplines, the phenomena that scientists study influence how their work is organized and carried out» (Hesse et al., 1993, s. 92). For å kunne forstå hvorfor forskere fra ulike fagdisipliner, kan ha veldig ulik datadelingspraksis, var det et behov for å se nærmere på mere fagspesifikke faktorer og hva som kan være kjennetegnende for de ulike fagdisiplinene. Becher & Trowler hevder at den komplekse naturen til en disiplin også kan forstås ut ifra et sosialt perspektiv, for eksempel gjennom graden av samarbeid og kommunikasjon forskerne imellom (Becher & Trowler 2001). Winnie Tam, fant i sin doktorgradsavhandling ut at hvordan forskere

samarbeider kan ha betydning for deres datapraksiser, - og spesielt datadelingspraksiser, for eksempel deres villighet til å dele, og med hvem de er villige til å dele (Tam, 2019, s. 101).

Jeg ser noen av de samme mønstrene som Winnie Tam fant i sin forskning. Tettere samarbeid mellom forskerne, og en større avhengighet av forskningsgruppen – gir ofte i utgangspunktet et annet rom for å dele forskningsdata.

I noen fagfelt kan det være et større behov for å dele data, derfor kan gjenbruk eller deling av data være en vanlig praksis i noen fagdisipliner, i andre disipliner kan det være mere unntaksvis at forskningsspørsmål vil kreve at man har behov for å hente inn data eller gjenbruke data fra andre. Hesse et al., påpeker blant annet at innenfor fysisk oseanografi samles det inn data fra avsidesliggende steder, noe som krever koordinering over lange avstander. Samarbeid og deling er nødvendig for å utføre fysisk oseanografisk forskning, siden ingen enkeltpersoner og få institusjoner har råd til å utføre det på egen hånd (Hesse et al., 1993). Disse dataene er også i mindre grad «eid» av enkeltpersoner som er avhengig av analyse og publisering av dataene for å fremme egen karriere (Zimmerman, 2003). Datadelingspraksiser kan også være sterkt påvirket av kulturelle og sosiale faktorer, som for eksempel i hvor stor grad forskerne kjenner på eierskap til dataene (Sterling, 1988, Sterling & Weinkam, 1990).

Ifølge litteraturen er ikke datadeling via databaser så vanlig, utenom for noen få fagdisipliner, og fordi data er så kontekstavhengige, kan de ofte være vanskelige å overføre på tvers av forskningsgrupper, lokasjoner og disipliner (Wallis et al., 2013, s. 3). Derfor er datadeling via databaser bare normen i noen få fagfelt (Borgman, 2015). Jeg har derfor i min studie valgt å ha en vid tilnærming til begrepet datadeling, og sett på ulike former for datadeling både innenfor og utenfor egen forskningsgruppe.

6.1.1. Datadelingspraksiser

Når jeg intervjuet forskerne, har jeg blant annet spurt om deres forskningsmetoder og hva slags datatyper forskningen deres genererer.

6.1.2. Geofysikk/geovitenskap

Innenfor geofysikk og geovitenskap, intervjuet jeg fire forskere med bakgrunn i fysisk oseanografi, og en innen paleoseanografi. Ved gjennomgang av intervjuene kommer det frem

at spesielt disse forskerne har en tradisjon for og praksis med å dele forskningsdata. Felles for disse forskerne er at de virker innen forskningsfelt som har funksjonelle måter å dele data på, og de deponerer gjerne dataene sine til ulike databaser eller datasentre. I fysisk oseanografi er det et kjennetegn at det finnes en del standardiserte og strukturerte data, og det er utviklet gode formater og metadatastandarder. Metadatastandarder brukes for å beskrive data på enhetlig måte. Metadata og metadatastandarder tilrettelegger for å formalisere og overføre informasjon. Fysisk oseanografi har de samme ligningene og metodene som meteorologi. Værvarsling og meteorologiske data er eksempel på data som karakteriseres som enkle og dette er data man har holdt på å standardisere i mer enn hundre år. En del av dataene innen fysisk oseanografi blir samlet inn via ulike selvregistrerende instrumenter. Samarbeid og deling av forskningsdata er nødvendig siden det er spesielt kostbart å utføre denne forskningen. De samarbeider ofte tett med beslektede fagdisipliner innenfor marinforskning og er avhengig av å dele data med hverandre. Innen fysisk oseanografi finnes det også en del data som er del av store datainnsamlinger. Der noe kan karakteriseres som stordata, og blir samlet inn av store team, og tilrettelagt for et stort forskningsmiljø. Stordata genereres ofte innenfor fagområder med internasjonalt etablerte standarder, avanserte infrastrukturer og administrative støttesystemer, til forskjell fra smådata som ikke har samme infrastruktur på plass.

6.1.3. Biovitenskap

Jeg har også intervjuet tre forskere innenfor biovitenskap. Der to forskere hadde bakgrunn innen marinbiologi. De samarbeider ofte med fysiske oseanografer, og sier det er stor forskjell på hva som kreves når det gjelder å håndtere ulike typer data, og at mange biologiske data ikke alltid er lett å dokumentere, sammenlignet med enkelte data innenfor fysisk oseanografi. Den ene forskeren innen marinbiologi peker på at den største utfordringen med noen av de biologiske dataene, er fravær av standarder for enkelte typer data, eller det kan være for mange standarder for andre typer data, slik at det ikke er enighet om hva som skal brukes. Det gjør det vanskelig både når man skal dele data og gjenbruke data. De to forskerne innen marinbiologi sier begge at det nok er en kultur innenfor deres fagdisiplin at man i liten grad deler forskningsdata via databaser eller lignende, men at det er noe som er i veldig endring, og det blir mer vanlig. Begge disse forskerne har imidlertid selv gjort data tilgjengelig i databaser,

men etterlyser bedre infrastruktur, mer kompetanse og opplæring, hvis de skal kunne tilgjengeliggjøre dataene sine på en god måte.

Den tredje forskeren innen biovitenskap jobber innenfor miljøtoksikologi – der man for eksempel ser på hvordan miljøgifter forflytter seg i økosystemet og samler seg opp i ulike dyr. Dataene fra denne forskningen blir lagret internt ved institusjonen, men det er ikke vanlig å sende dataene inn til databaser eller lignende. Når de publiserer artikler er det vanlig å skrive at data er tilgjengelig ved forespørsel, også fordi de ønsker å sikre at de kan ha en dialog med dem som skal bruke dataene, slik at de brukes på riktig måte. Denne forskeren uttrykte en viss skepsis mot å gjøre data offentlig tilgjengelig i databaser – og ønsker å dele data først og fremst som del av samarbeidsprosjekt, blant annet for å sikre at data blir gjenbrukt korrekt.

Ifølge Borgman utgjør økologiske data en motsats til klimadata som brukes i modeller. I forskning hvor det studeres klima og fysiske prosesser, finnes det oftere etablerte standardmålinger, mens forskning på biologiske organismer krever mer spesialiserte metoder og målinger (Borgman, 2010). Det kan derfor være spesielle utfordringer knyttet til å dokumentere disse dataene for gjenbruk (Zimmerman, 2003).

6.1.4. Samfunnsvitenskap/humaniora

For forskerne jeg intervjuet innenfor samfunnsvitenskap var det ikke vanlig å dele data utenfor egen forskningsgruppe, selv om den ene statsviteren unntaksvis har delt et kvantitativt datasett og kildemateriale som ikke er offentlig tilgjengelig. Den andre statsviteren har ikke mottatt forespørsler om å dele data. Det kan tyde på at det er mindre interesse for disse dataene i utgangspunktet, noe som kan virke inn på forskerens motivasjon for å dele data. I tillegg påpeker denne forskeren at de kvalitative dataene fra denne forskningen ofte er i et format som ikke er like egnet for deling, da det kan være biter av forskjellige publikasjoner og kilder, som ikke er samlet i tabeller man kan kopiere over. De tre forskerne fra samfunnsvitenskap brukte i stor grad kvalitative metoder. Alle tre brukte en del kildedata, og to av dem gjorde en del intervjuer. Kildedata eller annenhåndslitteratur er gjerne tilgjengelig i utgangspunktet, noe som gir mindre behov for å dele data. Data fra intervjuer er av etiske årsaker gjerne konfidensielle, og her setter sensitive data fra personintervjuer og hensyn til personvern store begrensinger for hva som kan deles. Begge forskerne innen statsvitenskap peker på at det ofte

er en betydelig grad av subjektivitet knyttet til de kvalitative dataene, og de mente at en del av disse dataene ikke egner seg for deling.

En forsker innen arkeologi – jobber med «klimaforskning og bevaringsstudier». Denne forskeren jobber i tett samarbeid med andre forskere fra flere ulike fagdisipliner. Forskeren forteller at dataene fra denne forskningen har størst verdi, når de deles med flest mulig – og at man trenger ulik fagkompetanse for å kunne tolke dataene korrekt. Måledataene eller de naturvitenskapelige dataene fra denne forskningen deles via en nettside. De kvalitative dataene fra den mere klassiske arkeologidelen, var det mere vanlig å holde tett til kroppen inntil forskningen var publisert. Imidlertid er det viktig å dele grunndataene fra denne forskningen, slik at andre kan omtolke og nytolke funnene. Datakilder fra denne forskningen blir gjerne gjenbrukt flere ganger, og blir gjerne mer verdifulle med tiden.

Samspillet med andre kan ha en sentral betydning for forskeres datadelingspraksiser. Egenskaper ved dataene kan ha betydning for hvilke sosiale praksiser som utformes rundt forskeres datadelingsadferd. Van House viser for eksempel til at sosiale praksiser oppstår rundt menneskers daglige aktiviteter og interaksjon med hverandre, og i samspill med materiale og andre ressurser (Van House, 2003).

Forskeres holdninger til datadeling, og deres faktiske datadelingspraksis er kompleks og avhengig av mange faktorer. Å dele data blir i stor grad sett på som positivt av de fleste innenfor denne studien, men noen møter også på barrierer som hindrer dem i å dele data mere vidt. I tillegg til dataenes karakter, og kulturelle normer innenfor de ulike fagdisiplinene, vil det også være eksterne og interne faktorer som kan virke inn på forskeres datadelingspraksiser.

6.1.5. Avgjørende faktorer for datadeling

Gjennomgang av analysen viser at det er variasjoner – så vel som fellestrekk – i forskeres holdninger, atferd og behov, i tilgjengelighet til infrastruktur, og i arten og effekten av policyer for datadeling, i de ulike disiplinene og fagområdene. Når jeg nå skal se nærmere på faktorer som kan ha innvirkning på forskeres datadelingspraksiser er det med tanke på deling av forskningsdata i databaser eller lignende.

Faktorer som kan virke inn på forskeres datadelingspraksiser er blant annet - tid og ressurser til rådighet, å tilgjengeliggjøre data kan være veldig tidkrevende. Om data skal være mulige å gjenbruke må de organiseres, tildeles metadata og dokumenteres. I noen datasentre kan man få hjelp til å sammenstille dataene, men det er ikke alle datatyper de tar imot. Og det er heller ikke alle databaser som tilbyr slik assistanse. Da blir det opp til forskeren å finne ut av det selv, og hvis det krever for mye, så er det lett å droppe det.

Noen av forskerne etterlyste mer kunnskap og ekspertise, for å kunne gjøre data tilgjengelig for andre. Forskerne innen fagområdene geofysikk og geovitenskap kan benytte seg av store sentraliserte datasentre, og kan dra nytte av ekspertise og ressurser innen datakurering som nok ikke alltid tilbys like konsekvent på lokalt eller nasjonalt nivå. Den ene marinbiologen peker blant annet på at den nasjonale infrastrukturen bør forbedres, for å kunne gjøre dataene gjenbrukbare og FAIR. Mangel på etablerte metadatastandarder kan også gjøre det vanskeligere å dele data. (Men det er nok heller ikke alle data som er like enkle å standardisere).

Å dele data kan også være et kostnads- og ressurs spørsmål. Flere forskere pekte på at det ikke er satt av midler til å gjøre den jobben som kreves for å tilgjengeliggjøre dataene for andre. Fordelen av en slik innsats går dessuten i stor grad til sekundærbrukeren.

En av de sterkeste innflytelsene på forskeres datadelingspraksiser er mandater og retningslinjer for datadeling. En årsak til den økte praksisen med å dele forskingsdata i databaser eller lignende, er utviklingen av et politisk rammeverk, som oppmuntrer eller krever tilgang til forskningsdata finansiert med offentlige midler. Flere av informantene spesielt fra geofysikk og geovitenskap forteller at de møter på krav om å tilgjengeliggjøre dataene sine, både fra finansører og fra tidsskrifter. Det settes som betingelse for finansiering og publisering at data skal gjøres tilgjengelig. Den ene marinbiologen forteller også at press fra finansører hvor det stilles strenge krav har gjort at man har speedet opp prosessen med å gjøre data tilgjengelig i databaser.

I gjennomgang av analysen fant jeg også ut at holdninger til å dele data kan påvirkes av oppfatningen om hvor «delbare» eller gjenbrukbare dataene vil være for andre. Det var for eksempel også noen forskere innen fysisk oseanografi som pekte på at ikke alle data egner seg for deling. Selv om dokumentasjon som metadata legges ved, kan det ofte være slik at kunnskapen som trengs for å forstå et datasett er taus. Det er ikke alltid mulig å uttrykke denne kunnskapen eksplisitt. Hvis forskeren ikke kan uttrykke eller dokumentere konteksten som trengs for å forstå et datasett, kan dataene også lettere mistolkes eller brukes utenfor sine forutsetninger. Derfor kan forskere være skeptiske til å legge ut slike data. Et eksempel som var nevnt var eksperimentelle data. Det var også flere barrierer med å dele kvalitative data.

Flere forskere peker også på at det mangler insentiver for å gjøre data tilgjengelig for andre. De blir belønnet for å lage en publikasjon, men verdien av å gjøre data tilgjengelig blir ikke tilstrekkelig verdsatt. Norsk publiseringssindikator, som er en del av grunnlaget for tildeling av midler til forskningsinstitusjoner gir hovedsakelig uttelling for publiserte artikler, og ikke dataene som ligger til grunn for dem (Forskningsrådet, 2021). Samtidig som at data kan være et essensielt grunnlag for annen innbringende forskningsaktivitet, og det ikke gir noen form for uttelling å dele dataene i seg selv, så kan dette skape en motsatt effekt – ved at man holder tilbake dataene.

Andre barrierer for å dele data inkluderer juridiske og etiske begrensninger, frykt for utnyttelse eller upassende bruk av dataene, og frykten for at deling av data kan sette fremtidige publiseringmuligheter i fare. Disse faktorene kombinert med den opplevde mangelen på karrierebelønninger for datadeling, utgjør sannsynligvis noen av årsakene til at deling og deponering av data i arkiv eller databaser ikke er mere vanlig, og derfor ofte er mer begrenset til fagdisipliner som har en sterk nytte av og behov for å dele data med hverandre i utgangspunktet.

Noen av forskerne viste til at de ville være motivert til å gjøre mer data tilgjengelig i fremtiden hvis de mottok ekstra midler for å dekke kostnadene ved utarbeidelse av data, hvis tilgjengeliggjøring av data ville styrke deres akademiske omdømme, hvis de visste hvordan

andre forskere brukte dataene deres, og hvis datadeling ble tatt i betraktning i fremtidige beslutninger om finansiering og karrierefremmende tiltak. I tillegg var det viktig å få anerkjennelse for arbeidet med å frembringe og dele datasett, og at data blir sitert ved gjenbruk. Hovedårsakene til å gjøre forskningsdata tilgjengelig for utenom finansierings- og tidsskriftskrav, er at det anses som god forskningspraksis og at det bidrar til å fremme forskningen i seg selv, og at det muliggjør sammenligning og validering av forskning. Noen av forskerne rapporterte om mer samarbeid og at sitering for publikasjoner der underliggende datasett er delt, blir god.

6.1.6. Holdninger

Forskerne var generelt positive til å dele data utenfor sin egen forskningsgruppe. Selv om forskerne fra samfunnsvitenskap viste til at det ligger noen begrensninger i å dele det datamaterialet som blir til i deres forskning. Mange av informantene delte oppfatningen om at data som er finansiert med offentlige midler er et kollektivt gode som flere skal kunne bruke. Det var likevel ikke alltid enige i detaljene rundt åpen tilgang og åpne data. Nesten samtlige informanter pekte på noen betingelser som må være på plass. Det som var nevnt flest ganger var førsteretten til å publisere sine forskningsresultater og mulighet for embargo. Noen nevnte at de må få tid til å kvalitetssikre dataene sine før det kreves at de skal tilgjengeliggjøres. Hvis ikke er det en fare for at det kan slå negativt ut for forskeren. Dersom dataene viser seg å ikke holde mål, kan man risikere sitt omdømme og fremtidige karrieremuligheter.

Flere av informantene uttrykte bekymring for at dataene kan bli kuppet før man selv rekker å publisere, spesielt gjelder dette dersom det er mer unike data eller banebrytende forskning. Sitter en på datamateriale, kan man også sitte på kapital for nye søknader og for mulige publikasjoner. De som har samlet inn data kan ha investert betydelig arbeidsinnsats og kreativitet i planlegging og innsamling, og derfor kjenne på et eierskap og et ansvar for dataene sine. Forskere kan derfor ofte uttrykke en ambivalens, på den ene siden støtter de idéen om åpen forskning og åpne data, på den andre siden må dette balanseres med å ikke ofre egne karrieremuligheter.

I gjennomgang av analysen ser det ut til at både datadeling og datagjenbruk er regulert av mange uoffisielle fellesskapsnormer. Noe som ble nevnt av mine informanter var at de kan

kjenne på forpliktelser overfor de som har delt data. Respekt for andres arbeid, kan gjøre at man ikke bare «forsyner» seg av data som er deponert til databaser eller arkiv, men at det opprettes en form for samarbeid eller medforfatterskap. Det kan også være man har behov for dataskaperens kompetanse for å kunne gjenbruke forskningsdataene. Derfor ser deling av forskningsdata ofte ut til å innebære en form for samarbeid, også for data som ligger ute i arkiv – dersom de er samlet inn av enkeltforskere.

Dette kan stå i kontrast til det inntrykket man får når man leser ulike retningslinjer for deling av forskningsdata. Som man kanskje kan si innebærer et dekontekstualisert syn – der data står som noe løsrevet og uavhengig av forskerne som frembragte dem. Retningslinjer for datadeling ser derfor ikke ut til å anerkjenne datadeling som den relasjonsrelaterte og sosiale praksisen det ofte er å dele data, men utgår gjerne fra et teknisk eller teknologisk løsningsorientert perspektiv på utfordringene rundt datadeling.

6.2. Avsluttende diskusjon

Denne studien har undersøkt datadelingspraksiser til et lite utvalg forskere ved norske forskningsinstitusjoner, og hvilke erfaringer og holdninger de har til å dele dataene sine med andre. I intervjuene med forskerne har jeg spurt om deres forskningsmetoder, datatyper, hvordan de lagrer og håndterer dataene. I tillegg har jeg vært innom temaer som samarbeid, kultur, policyer for datadeling, samt motiv og barrierer for å dele data. Spørsmålene og temaene er viktige å studere ut ifra et samfunnsperspektiv ettersom det gjøres betydelige investeringer relatert til å gjøre forskningen mer åpen, fordi man anser at det vil gagne både samfunnet og forskningen i seg selv. Datadeling gjør det mulig å bygge videre på arbeidet til andre. Det er med dette i tankene at forskningsfinansiører, forlag og myndigheter har begynt å oppfordre eller til og med kreve at data fra offentlig finansiert forskning skal gjøres tilgjengelig. Allikevel er ikke datadeling noen ny foreteelse i forskersammenheng, men de eksterne kravene enkeltforskere møter for å gjøre dem tilgjengelig for andre er det (Mauthner & Parry, 2013).

Åpen forskning kan sies å være karakterisert av en «top down» policy, som kan påvirke eller endre måten forskning utføres på. Tidligere forskning viser at det er et gap mellom ønskene om

åpenhet rundt deling av forskningsdata, og den faktiske delingen som blir gjort (Tenopir 2011, Wallis et. al., 2013).

Å forstå hvorfor det kan være slik krever et nærmere blikk på datapraksisen til forskerne selv, og å ta i betraktning den samlede forskningsprosessen, med dens oppgaver, verktøy og arbeidsflyter. I tillegg til å se nærmere på hva data kan bety for dem.

Data kan sies å være den primære eiendelen for en forsker – og betegnes derfor gjerne som forskerens kapital. Krav om å frigi data på publiseringstidspunktet forutsetter at forfatterne er «ferdige» med datasettet (Borgman, 2015, s. 217). I tilfeller der en forskerkarriere er basert på en mer langsiktig studie av for eksempel spesifikke arter, eller et sett med artefakter, blir data mer verdifulle ettersom de akkumuleres. Disse forskerne kan være motvillige til å tilgjengeliggjøre data assosiert med en spesifikk publikasjon fordi det kan bety å frigjøre mange års arbeid. Med mindre forskere får noe igjen for det kan det være vanskelig å overbevise dem om å gi slipp på disse verdiene, spesielt hvis det er involvert store anstrengelser og kostnader i innsamlingen av dem (Borgman, 2007, Edwards et al, 2011). Konkurransen ære og berømmelse, samt forskningsmidler kan være intens og derfor kan det være motvilje mot å dele data, fordi det kan bety å gi slipp på fremtidige avkastninger, i form av for eksempel publikasjoner eller renommé (med tilhørende fordeler) (Birnholtz & Bietz, 2003). Dersom data deles, - kan noen andre komme en i forkjøpet og gjøre det en selv hadde tenkt å gjøre. Å dele disse dataene kan bety å risikere å tape disse fordelene. Ifølge Borgman kan datasett også være verdier som blir brukt til å forhandle med andre forskere, brukt som utveksling i samarbeid og brakt med som medgift. Forskere kan også beskytte samarbeid og arbeidsinvesteringer ved å ikke tilgjengeliggjøre data tidligere enn nødvendig (Borgman, 2015).

Avhengig av hva slags type data man har, kan data også være ansvar og forpliktelser. Å forvalte dem kan kreve menneskelige ressurser. Bevaring av data kan stille ressurskrav til både langtidslagring og datakuratering (Borgman, 2015). Tid brukt på å tilgjengeliggjøre data, er tid som ikke blir brukt på andre forskningsaktiviteter. Det er heller ikke alltid klart hvem som har eierskapet til dataene. Spenninger rundt eierskap og kontroll av data kan også påvirke datadeling og gjenbruk.

6.2.1. Sosioteknisk perspektiv

Ved å benytte et sosioteknisk perspektiv, kan man se nærmere på tekniske og sosiale aspekter som kan ha innvirkning på forskeres datadeling. Der hverken tekniske eller sosiale aspekter settes over hverandre, men ses som gjensidig avhengig av hverandre. Datadelingspraksisen eller atferden kan bestemmes av hvor godt disse delene passer sammen eller samspillet mellom sosiale og tekniske aspekter. Siden det må tas hensyn til teknologien, så vel som til uformelle og formelle interaksjoner mellom mennesker.

Hvis vi legger til grunn at forskning er et sosialt foretak, må vi også spørre hvordan data er involvert i den sosiale verden. For dette trekkes det veksler på tidligere teori om at kunnskap er sosialt konstruert og teorien om praksisfellesskap (Lave & Wenger), der læring er tett sammenkoblet med sosiale praksiser, og materielle fenomen. Forskningsdata er bearbeidet for spesifikke forskningsspørsmål, og er lokale og historisk situerte artefakter (Gitelman, 2013, Star & Griesemer, 1989). Data er ikke naturlige objekter med en egen essens eller mening i seg selv, de eksistere i en kontekst. Kunnskapsoverføring i dette tilfellet er ikke kun et spørsmål om å dele et sett med instruksjoner, men er i høy grad en sosial prosess som består av å lære seg praksiser som ikke alltid er lett å dokumentere. Data kan i mange tilfeller ikke bli overlevert enkelt fra en forsker til en annen.

Datadelingssystemer har utviklet en rekke tilnærminger for å forsøke å løse disse utfordringene. For eksempel metadatastandarder og FAIR-prinsipper som skal gjøre det enklere å overføre data og informasjon mellom ulike personer. Metadata og annen dokumentasjon kan gi kontekstinformasjon om hvordan dataene er frembragt eller innsamlet, og hvordan de er behandlet og prosessert. Mens disse tilnærmingene forsøker å ta hånd om eksplisitte teknologiske utfordringer iboende i datadeling, så er det ikke klart at de i tilstrekkelig grad tar hånd om de tause eller sosiale utfordringene ved datadeling. For eksempel kan metadata gi informasjon om dataene, men de kan også avhenge av en god del fagekspertise for å kunne tolkes og forstås. I noen tilfeller kommer kontekstinformasjonen forskerne trenger fra dokumentasjon, i andre tilfeller blir den utvekslet som en del av et forskerfellesskap (Faniel & Jacobsen, 2010).

Datadeling på forespørsel forskere imellom kan være veldig effektivt fordi forskere kan diskutere innhold, kontekst, styrker og svakheter og anvendelsen av et spesifikt datasett til et fenomen (Borgman, 2015). Tilgangsprotokoller kan sette opp en mur rundt et datasett, men de kan ikke i tilstrekkelig grad sørge for de subtile sosiale realitetene ved å skaffe tilgang til en usynlig kollega. Noen ganger er man avhengig av sosial interaksjon – og en annens forskers fagkunnskap og ekspertise (Faniel & Jacobsen, 2010).

Det kan også være andre sosiale barrierer for å effektivt kunne bruke disse delte dataene i praksis. Når data som tidligere var privat blir offentlig tilgjengelig reiser det seg en rekke spørsmål, som for eksempel frykt for at data skal misbrukes eller feiltolkes og gjenbrukes vanskeligheter med å bedømme kvaliteten på dataene. Disse potensielle utfordringene kan oppstå når data skal krysse grenser (boundaries) (Star 1989, Star & Griesemer, 1989). Ved at man gjør data tilgjengelig til og fra folk i andre profesjoner eller praksisfellesskap, og ved å gjøre offentlig det som tidligere var privat (Van House, 2004). Delingen av data har tidligere gjerne funnet sted innenfor en sosial, organisatorisk eller profesjonsbasert avgrensning, som har bidratt til å sikre en adekvat forståelse og anvendelse, samt gjensidig utnyttelse av dataene. Deling av forskningsdata i åpne databaser/dataarkiv har potensial for dekontekstualisering av informasjon og å gjøre det mer tilgjengelig utenfor den sosiale kretsen som det var produsert innenfor og som det ble produsert for (Van House, 2003, s. 277).

Å bruke data samlet inn av andre er en tillitshandling. Gjenbruk av datasett fordrer at brukeren vurderer kvaliteten og påliteligheten av innholdet. Å vurdere påliteligheten til slik informasjon, er ikke alltid lett (Van House, 2003, s. 278). Forskere trenger også å vite noe om kvaliteten på dataene de vurderer å gjenbruke, og om dets opprinnelige formål med dataene er i samsvar eller kompatibelt med tiltenkt bruk av dem.

Golden påpeker at teknologi som er tenkt å forbedre et system, også kan gjøre at andre deler av systemet kan gå tapt, og med det ønskelige funksjoner (Golden, 2013). For eksempel påpekte en av informantene i denne studien at man ikke nødvendigvis kjenner til kvaliteten på de datasettene som ligger ute i arkiv. Det ble påpekt at forskningsresultater utgitt i en vitenskapelig publikasjon har gjennomgått en kvalitetssjekk og en fagfellevurdering, mens et

datasett er en haug med målinger og de kan være feil. Elektronisk publisering av data kan gi bekymringer om upålitelige data, og dataenes kvalitet. Van House viser til lignende eksempler på misforholdet mellom den nye teknologien, og tidligere praksiser og institusjonelle arrangementer, som fremhever de komplekse gjensidige avhengighetene til sosioteknisk nettverk (Van House, 2004, s. 61). Mange ulike valg kan være tatt når man samler inn data. Når vi skal forstå hva som skal til for å kunne gjenbruke et datasett, så må vi forstå hvordan folk bestemmer seg for å stole på andres datasett, og hvordan de inkorporerer dette i sitt eget arbeid eller forskning, og hvordan deling av data i databaser kan støtte eller underminere disse prosessene. Strategiske investeringer i kunnskapsinfrastrukturer, slik som databaser, menneskelige ressurser med ekspertise i dataforvaltning, bedre verktøy, og metoder som sørger for kreditering ved datadeling kan være med på å øke tilgjengeliggjøring og bruk av forskningsdata.

Det kan være sosiale årsaker til barrierer for datadeling som teknologiske systemer i den nærmeste fremtid ikke vil kunne løse. På samme tid kan det også være muligheter for å tilrettelegge for teknologiske systemer som støtter opp under sosiale relasjoner og sosial atferd og som kan bidra til mer deling av forskningsdata. Som blant annet denne studien viser, finnes det på noen fagområder som fysisk oseanografi - systemer og infrastruktur som ser ut til å være rimelig effektive for å dele forskningsdata. Et kjennetegn er at de har stor grad av omforente standarder og stor grad av enighet om metoder og hvordan man løser ulike typer utfordringer. Problemene som tas opp er tilstrekkelig store til at forskere er avhengig av store grupper for å kunne håndtere dem effektivt. I motsetning til forskning som gjøres i mindre grupper (i den lange halen av forskning), og som har et videre utvalg av spørsmål, metoder og dataformater, der denne variasjonen kan være med på å komplisere datadeling i databaser.

Retningslinjer for datadeling forutsetter at forskningsdata er nyttige for andre, og at andre vil gjenbruke dem (European Commission, 2020, Wilkinson et al., 2016). Det egentlige målet er ikke deling av data, men gjenbruk av data. Mer deling av offentlig finansierte forskningsdata er en ambisjon, men mer deling i seg selv vil ikke nødvendigvis føre til mer gjenbruk. En voksende mengde samfunnsvitenskapelig forskning viser til at deling av data er en kompleks sosioteknisk

prosess, som gjør det vanskelig å forutsi av hvem, når, hvordan, hvorfor og om forskningsdata vil bli gjenbrukt (Borgman, 2015, Mosconi et al., 2019).

6.3. Refleksjon og begrensing ved studien

Jeg valgte kvalitativ metode for å få svar på oppgavens problemstilling og forskningsspørsmål. Valget av metode har både fordeler og ulemper. Kvalitative metoder er som regel ressurskrevende og setter naturlige begrensinger for størrelsen på utvalget. Bruk av kvalitativ metode for datainnsamling og analyse har gitt innsikt i forskeres datadelingspraksiser, erfaringer og holdninger til datadeling, som ikke ville ha kommet like godt frem ved hjelp av kvantitativ metode. Det er likevel flere begrensinger ved min studie. For det første er antallet deltagere i min studie liten. Dette er typisk for bruk av kvalitativ metode, men det begrenser generaliserbarheten av funn, og gjør det vanskelig å vurdere i hvilken grad min studie er representativ. Beslutningen om å intervjuere forskere fra flere ulike fagdisipliner, og med ulike typer av data, baserte seg på et ønske om å forsøke å få med ulike perspektiver på deling av forskningsdata. Konsekvensen er at det er vanskelig å oppnå metning i intervjuene når det gjelder for eksempel spørsmål som dreier seg om spesifikke datadelingspraksiser og faktorer som kan virke inn på deling av forskningsdata. Det kan være at denne studien inneholder hull og skjevheter som skyldes det begrensede utvalget, og at synspunkter eller meninger som kommer frem i min studie ikke er representativt for andre forskere innenfor den samme fagdisiplin som er representert her. I etterpåklokskapens lys ville det nok vært mer interessant og fruktbart med et utvalg bestående av informanter fra en eller to forskningsdisipliner. Det ville muliggjort sammenligninger og generalisering på en helt annen måte. Det hadde også vært ønskelig å komplettere intervjuene med observasjoner på plass i forskeres arbeidsmiljø, noe som dessverre ikke var mulig å gjennomføre innenfor rammen av dette masteroppgaveprosjektet.

6.4. Forslag til videre forskning

Datadelingsretningslinjer og offentlig politikk beskriver sjelden data som det komplekse og heterogene fenomenet det faktisk er, og kan i stor grad ignorere ulikheter i datapraksiser mellom de enkelte fagdisiplinene. Dessuten heller forenkler enn å anerkjenne de mange hindringene som finnes før en realisering av disse kravene er mulig. Mauthner og Parry (2013)

tar opp de sosiale forholdene i relasjon til datadeling, og påpeker at et problem med dagens datapolicyer er at de først og fremst behandler de tekniske problemene i selve datadelingsinfrastrukturen, og går glipp av de relasjonsbaserte årsakene. Derfor bør man i tillegg til å se på forbedringer i den tekniske infrastrukturen for datadeling, også vektlegge en bedre forståelse av de mer relasjonsbaserte og sosiale faktorene ved datadeling blant forskere.

Deling av forskningsdata og gjenbruk av forskningsdata er to tema som ofte behandles hver for seg. De er komplekse nok å forstå hver for seg, men de er innbyrdes forbundet med hverandre. For at data skal kunne gjenbrukes av andre enn dem som frembragte dem, må de først deles. I politiske dokumenter og policyer for forskningsdata, snakkes det mye om verdien av å dele forskningsdata, mens det er lite som omhandler hvordan gjenbruk av data faktisk skal foregå. Hva impliserer det i praksis, og hva kreves det av dem som skal gjenbruke forskningsdata? Retningslinjer for datadeling bør også basere seg på kunnskap om hvordan forskere gjenbraker data som har blitt delt med dem. Vi vet lite om hvordan brukere lokaliserer data, hvordan de integrerer dem i sin forskning, hvilke hindringer de møter, og hvordan data blir brukt til å teste nye hypoteser eller skape ny informasjon.

Etter hvert som metodene for datainnsamling blir mer avanserte, vil forskerne generere stadig større mengder data. Viktigheten av å lokalisere og identifisere relevante data, og forstå og stole på dem vil øke sammen med nødvendigheten av tilstrekkelig dokumentasjon. Under slike omstendigheter vil forståelsen av gjenbruksatferd, spesielt kontekstinformasjonen som trengs og strategiene og ressursene som brukes for å vurdere kvaliteten på dataene øke i betydning. Å forstå hva slags kontekst gjenbrukere av datasett trenger, og ikke bare fra perspektivet av hva de som produserer eller skaper data kan yte eller bidra med, kan være nyttig for utviklingen av databaser/dataarkiv og tilhørende infrastruktur. Beslutninger om nivå av metadatainformasjon for dokumentasjon hadde vært lettere hvis man visste mer om mulige gjenbrukere og bruk av forskningsdata. For eksempel, hvilke standarder vil være viktig. Hvordan blir metadata brukt? Og hvordan bedømmes datakvalitet? Et viktig spørsmål for vitenskapen og for offentlig politikk er å spørre hva slags datagjenbruk som blir muliggjort ved tilgang til offentlige dataarkiver og hva slags typer som ikke muliggjør gjenbruk. Det hadde vært interessant å se flere kvalitative

studier omkring hvordan forskere stiller seg til å anvende seg av andres innsamlede data, samt hva slags påvirkning dette kan ha på forskningen.

Studier omkring forskeres datadelingspraksiser og deres holdninger til og erfaringer med både deling og gjenbruk av forskningsdata vil være av stor betydning, for at forskningsinstitusjoner skal kunne tilby riktig støtte og bygge opp infrastrukturer knyttet til ulike faser av forskningsdatahåndteringen.

7. Litteraturliste

- Ajzen, I. (2005). *Attitudes, Personality and Behavior*. (2nd. Ed.) Berkshire: McGraw-Hill Education.
- Aksnes, D. W. & Rørstad, K. (2015). *Norsk polarforskning – forskning på Svalbard: ressursinnsats og vitenskapelig publisering – indikatorer 2014*. (Rapport 2015:37). Oslo: NIFU.
- Baker, K. S. & Yarmey, L. (2009). Data Stewardship: Environmental Data Curation and a Web-of-Repositories. *International Journal of Digital Curation*, 4 (2), 12-27. DOI: <https://doi.org/10.2218/ijdc.v4i2.90>
- Baker, K. S., Duerr, R. E. and Parsons, M. A. (2015). Scientific Knowledge Mobilization: Co-evolution of Data Products and Designated Communities. *International Journal of Digital Curation*, 10 (2), 110-135. <https://doi.org/10.2218/ijdc.v10i2.346>
- Becher, T. & Trowler, P. R. (2001). *Academic tribes and territories: intellectual enquiry and the culture of disciplines*. Philadelphia, PA: Open University Press.
- Berger, P. L. & Luckmann, T. (2000). *Den samfunnsskapte virkelighet*. Bergen: Fagbokforlaget.
- Birnholtz, J. P., & Bietz, M. (2003). Data at work: Supporting sharing in science and engineering. *ACM 2003 International Conference on Supporting Group Work, Sanibel Island, USA*, 339-348.
- Borgman, C.L. (2007). *Scholarship in the Digital Age: Information, Infrastructure and, the Internet*. Cambridge, MA: MIT Press.
- Borgman, C. L. (2010). Research Data: Who Will Share What, with Whom, When, and Why? *RatSWD Working Paper No. 161*. <http://dx.doi.org/10.2139/ssrn.1714427>
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63 (6), 1059-1078. <https://doi.org/10.1002/asi.22634>
- Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge, MA: MIT Press.
- Borgman, C. L., Darch, P. T., Sands, A. E., & Golshan, M. S. (2016). The Durability and Fragility of Knowledge Infrastructures: Lessons Learned from Astronomy. I: *Proceedings of the Association for Information Science and Technology*. 53 (1), 1–10. Copenhagen. <https://doi.org/10.1002/pra2.2016.14505301057>

- Borgman, C. L., Scharnhorst, A. and Golshan, M.S. (2019), Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70 (8), 888-904. <https://doi.org/10.1002/asi.24172>
- Bowker, G. C. (2000). Biodiversity datadiversity. *Social studies of Science*, 30 (5), 643-683. <https://doi.org/10.1177/030631200030005001>
- Bowker, G. C. (2005). *Memory practices in the Sciences*. Cambridge, Mass: MIT Press.
- Bowker, G. C., Baker, K., Millerand, F., Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. I J. Hunsinger, L. Klastrup, & M. Allen, (Red.) *International Handbook of Internet Research* (s.97-117). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-9789-8_5
- Bryman, A. (2008). *Social Research Methods* (3rd ed.). Oxford: Oxford University Press.
- Carlson, S. & Anderson, B. (2007) What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, 12 (2), 635-651. <https://doi.org/10.1111/j.1083-6101.2007.00342.x>
- Creswell, J. W. (2013). *Qualitative Inquiry & Research design: Choosing among Five Approaches* (3rd ed.). Los Angeles: Sage Publications.
- Dalland, O. (2020). *Metode og oppgaveskriving* (7. utg.). Oslo: Gyldendal akademisk.
- DAMVAD. (2014). *Sharing and archiving of publicly funded research data*. Report to the Research Council of Norway. København: DAMVAD.
- Davenport, T. & Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know*. Boston: Harvard Business School Press.
- Edwards, P. N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press
- Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013) Knowledge Infrastructures: Intellectual Frameworks and Research Challenges. Ann Arbor: Deep Blue. <http://hdl.handle.net/2027.42/97552>
- European Commission. (2016). Open Innovation, Open Science, Open to the world – A Vision for Europe. DG Research and Innovation. <https://doi.org/10.2777/061652>
- European Commission. (2018). Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information. C/2018/2375. Hentet fra <http://data.europa.eu/eli/reco/2018/790/oj>

- European Commission, Directorate-General for Research and Innovation, Mendez, E. (2020). *Progress on open science: towards a shared research knowledge system: final report of the open science policy platform*, (R, Lawrence, editor) Publications Office.
<https://data.europa.eu/doi/10.2777/00139>
- Faniel, I. M. (2009). Unrealized potential: The socio-technical challenges of a large scale cyberinfrastructure initiative. *University of Michigan, School of Information, Ann Arbor*. Hentet fra: <https://hdl.handle.net/2027.42/61845>
- Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work* 19, 355–375. <https://doi.org/10.1007/s10606-010-9117-8>
- Forskningsrådet. (2017). *Tilgjengeliggjøring av forskningsdata: Policy for Norges forskningsråd*. Oslo: Norges forskningsråd. Hentet fra <https://www.forskningsradet.no/siteassets/publikasjoner/1254032061080.pdf>
- Forskningsrådet. (2019). *Forskningsrådets policy for åpen forskning: Gjeldende fra 2020*. Oslo: Norges forskningsråd. Hentet fra <https://www.forskningsradet.no/contentassets/98222e889c004c108a83bb4d32816779/policy-åpen-forskning.pdf>
- Forskningsrådet. (2020). *Evalueringen av norsk polarforskning: Oppfølgingsplan*. Lysaker: Norges forskningsråd.
- Forskningsrådet. (2021). *Hvordan skal vi dele forskningsdata? Utredning og anbefalinger om lisensiering og tilgjengeliggjøring*. Oslo: Norges forskningsråd.
- Fry, J., Lockyer, S., Oppenheim, C., Houghton, J. Rasmussen, B. (2008). Identifying benefits arising from the curation and open sharing of research data produced by UK Higher Education and research institutes. Loughborough University. Report.
<https://hdl.handle.net/2134/4600>
- Genova, F. & Horstmann, G. (2016). Long tail of data: e-IRG task force report, e-Infrastructure Reflection Group. Hentet 12. november 2021 fra <https://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf>
- Gillies, V. and Edwards, R. (2005). Secondary analysis in exploring family and social change: addressing the issue of context, *Forum: Qualitative Social Research*, 6(1).
<http://www.qualitative-research.net/index.php/fqs/article/view/500/1076>
- Gitelman, L. (Ed.). (2013). *Raw Data Is an Oxymoron*. Cambridge, MA: MIT Press
- Golden, T. D. (2013). Sociotechnical Theory. I Kessler E. H. (red.) *Encyclopedia of Management Theory*. Thousand Oaks: SAGE Publications, Ltd. 752-755.

- Goodwin, J. (2012). *Sage Secondary Analysis*. London: Sage
- Gottshcalc, P. (2003). *Ledelse av intellektuell kapital*. Oslo: Universitetsforlaget.
- Hesse, B. W., Sproull, L. S., Kiesler, S. B., & Walsh, J. P. (1993). Returns to science: Computer network in oceanography. *Communications of the ACM*, 36 (8), 90-101.
<https://doi.org/10.1145/163381.163409>
- Hofstad, K. (2020). Fysisk konstant. I *Store norske leksikon*. Hentet 12. november 2021 fra https://snl.no/fysisk_konstant
- Høyland, B. & Hegre, H. (2008). Ny redaksjon i norsk statsvitenskapelig tidsskrift. *Norsk statsvitenskapelig tidsskrift*, 24 (3-5), 3-5. https://www.idunn.no/nst/2008/01-02/ny_redaksjon_i_norsk_statsvitenskapelig_tidsskrift
- Johannessen, A. Tufte, P.A., Christoffersen, L. (2016). *Introduksjon til samfunnsvitenskapelig metode* (5. utg.). Oslo: Abstrakt.
- Kennan, M. A. (2017). 'In the eye of the beholder': Knowledge and skills requirements for data professionals. *Information Research: an international electronic journal*, 22 (4), Artikkel 1601. <http://www.informationr.net/ir/22-4/rails/rails1601.html>
- Kim, Y. & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: the roles of attitudes, norms and repositories, *Library and Information Science Research*, 37 (3), 189-200. <https://doi.org/10.1016/j.lisr.2015.04.006>
- Kirchner, S. R. (1997). Focus on: database integration and management for call centers. *Telemarketing*, 16 (2), 22-24.
- Knorr-Cetina, K. D. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Kowalczyk, S., & Shankar, K. (2013). Data sharing in the sciences. *Annual Review of Information Science and Technology*, 45(1), 247-294
<https://doi.org/10.1002/aris.2011.1440450113>
- Kunnskapsdepartementet. (2017). *Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata*. Hentet fra <https://www.regjeringen.no/contentassets/3a0ceea1c9b4611a1b86fc5616abde7/no/pdf/f-4442-b-nasjonal-strategi.pdf>
- Kuipers, T. & Hoeven, J. van der (2009). *Insight into digital preservation of research output in Europe*. PARSE.Insight. Hentet fra <https://libereurope.eu/wp->

<content/uploads/2010/01/PARSE.Insight.-Deliverable-D3.4-Survey-Report.-of-research-output-Europe-Title-of-Deliverable-Survey-Report.pdf>

- Kurata, K., Matsubayashi, M. and Mine, S. (2017). Identifying the complex position of research data and data sharing among researchers in natural science, *Sage Open*, 7 (3) 1-13, <https://journals.sagepub.com/doi/10.1177/2158244017717301>
- Kvale, L. H. (2012). *Data Sharing in the Life Sciences: A Study of Researchers at The Norwegian University of Life Sciences* (Mastergradsavhandling). Høgskolen i Oslo og Akershus. Hentet fra <http://hdl.handle.net/10642/1269>
- Kvale, S. & Brinkmann, S. (2015). *Det kvalitative forskningsintervju* (3. utg.). Oslo: Gyldendal.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
- Leonelli, S. (2010). Packaging small facts for re-use: Databases in model organism biology. I M. Morgan & P. Howlett (Red.). *How well do facts travel? The dissemination of reliable knowledge* (s. 325-348). Cambridge: Cambridge University Press.
- Mauthner, N. S. & Parry, O. (2013). Open Access Digital Data Sharing: Principles, Policies and Practices, *Social Epistemology*, 27 (1), 47-67, <https://doi.org/10.1080/02691728.2012.760663>
- Meld. St. 25 (2016–2017). *Humaniora i Norge*. <https://www.regjeringen.no/no/dokumenter/meld.-st.-25-20162017/id2545646/sec1>
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago press.
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., Da Silva Santos, L. O. B., Wilkinson, Mark D. (2017). Cloudy, Increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49-56. DOI:10.3233/isu-170824
- Moore, N. (2007). (Re)Using Qualitative Data? *Sociological Research Online*, 12 (3). <https://doi.org/10.5153%2Fsro.1496>
- Mosconi, G., Li, Q., Randall, D., Karasti, H., Tolmie, P., Barutzky, J., Korn, M., Volkmar, P. (2019). Three Gaps in Opening Science. *Computer Supported Cooperative Work*, 28 (3-4), 749–789. <https://doi.org/10.1007/s10606-019-09354-z>

NISO. (2004). *Understanding Metadata*. Bethesda, MD: NISO.

Organisation for Economic Co-operation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris: Organisation for Economic Co-Operation and Development.
<https://www.oecd.org/sti/inno/38500813.pdf>

Organisation for Economic Co-operation and Development. (2015). "Making Open Science a Reality", *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris <http://dx.doi.org/10.1787/5jrs2f963zs1-en>

Page, A. G. og Moen, E. T. (2020). En syklisk vei til profesjonell kunnskap: Praksisperiodens påvirkning på kunnskapsproduksjon i lærerutdanning. I A. Løhre & A. B. Lund (Red.), *Studenten skal bli lærer: Kunnskap, identitet og profesjonsutvikling* (Kap. 5, s. 105–118). Oslo: Cappelen Damm Akademisk. <https://doi.org/10.23865/noasp.98.ch5>

Pasquetto, I. V., Sands, A. E., Darch, P. T., & Borgman, C. L. (2016). Open Data in Scientific Settings: From Policy to Practice. I: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1585–1596. New York, USA: ACM.
<https://doi.org/10.1145/2858036.2858543>

Pasquetto, I. V., Randles, B.M. and Borgman, C.L. (2017). On the Reuse of Scientific Data. *Data Science Journal*, 16 (8), 1-9. <http://doi.org/10.5334/dsj-2017-008>

Pasquetto, I. V. (2018). PhD Dissertation - From Open Data to Knowledge Production: Biomedical Data Sharing and Unpredictable Data Reuses. UCLA: Center for Knowledge Infrastructures. Hentet fra: <https://escholarship.org/uc/item/1sx7v7>

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review*, 1(2), 1-34.
<https://doi.org/10.1162/99608f92.fc14bf2d>

Polanyi, M. (1966). *The tacit dimension*. Chicago: The University of Chicago Press.

Pryor, G. (2009). Multi-scale data sharing in the life sciences: some lessons for policy makers. *The International Journal Digital Curation*, 4 (3), 71-82.
<https://doi.org/10.2218/ijdc.v4i3.115>

Repstad, P. (2007). *Mellom nærhet og distanse: kvalitative metoder i samfunnsfag*. (4. utg.). Oslo: Universitetsforlaget

- Ringdal, K. (2013). *Enhet og mangfold: samfunnsvitenskapelig forskning og kvantitativ metode*. (3.utg.) Bergen: Fagbokforlaget
- Ringdal, K. (2018). *Enhet og mangfold: samfunnsvitenskapelig forskning og kvantitativ metode*. (4. utg.). Bergen: Fagbokforlaget
- Ryle, G. (1949). *The concept of mind*. (60th Anniversary Edition). New York: Routledge.
- Skagen, T. (2015) *Forskningsdata på Handelshøyskolen BI - Hva er forskernes behov?* Høgskolen i Oslo og Akershus. Hentet fra: <http://hdl.handle.net/10642/2707>
- Smeby, J.-C. (2001). Kunnskapens kognitive og sosiale strukturer: En teoretisk og empirisk studie av fagforskjeller ved universitetene. NIFU-rapport 2001:08. <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/274148/NIFUrapport2001-8.pdf?sequence=1>
- Star, S. L. (1995). *Ecologies of Knowledge: Work and Politics in Science and Technology*. Albany, NY: State University of New York Press.
- Star, S. L. & Griesemer, J. (1989). Institutional ecology, «translations», and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-1939, *Social studies of Science*, 19, 387-420.
- Sterling, T. D. (1988). Analysis and reanalysis of shared scientific data. *Annals of the American Academy of Political and Social Science*, 495, 49-60.
- Sterling, T. D., & Weinkam, J. J. (1990). Sharing scientific data. *Communications of the ACM*, 33(8), 113-119.
- Swan, A. & Brown, S. (2008) To share or not to share: Publication and quality assurance of research data outputs. A report commissioned by the Research Information Network (RIN). Hentet fra <http://eprints.soton.ac.uk/id/eprint/266742>
- Tam, W., Fry, J. & Proberts, S. (2014). The disciplinary shaping of research data management practices. In *iConference 2014 Proceedings*, 721-728. <https://doi.org/10.9776/14338>
- Tam, W. T. (2016). *Discipline and research data in geography*. Loughborough University. Thesis. <https://hdl.handle.net/2134/26997>
- Tenopir, C, Allered S., Douglass K, Aydinoglu A. U., Wu L, Read, E., et al. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6(6): e21101. <https://doi.org/10.1371/journal.pone.0021101>

- Tenopir, C., Dalton, E.D., Allard, S., Frame M, Pjesivac I, Birch B, ... Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE* 10(8), 1-24. <https://doi.org/10.1371/journal.pone.0134826>
- Tenopir, C., Christian, L., Allard, S., & Borycz, J. (2018). Research data sharing: Practices and attitudes of geophysicists. *Earth and Space Science*, 5, 891– 902. <https://doi.org/10.1029/2018EA000461>
- Thoegersen, J. L. & Borlund, P. (2022). Researcher attitudes toward data sharing in public data repositories: a meta-evaluation of studies on researcher data sharing, *Journal of Documentation*, 78 (7), 1-17. <https://doi.org/10.1108/JD-01-2021-0015>
- Tjora, A. (2021). Sosialkonstruktivisme. I Store norske leksikon. Hentet 9. september 2021 fra <https://snl.no/sosialkonstruktivisme>
- Van de Sandt, S., Dallmeier-Tiessen, S., Lavasa, A. and Petras, V., 2019. The Definition of Reuse. *Data Science Journal*, 18(1), 1-19. <http://doi.org/10.5334/dsj-2019-022>
- Van den Eynden, V. & Bishop, L. (2014). Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective. UK Data Archive. (Knowledge Exchange Report). https://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf
- Van den Eynden, V., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon D., et al. (2016) Survey of Wellcome researchers and their attitudes to open research. Wellcome Trust. Technical Report. <https://doi.org/10.6084/m9.figshare.4055448.v1>
- Van House, N. A., Butler M. H., Schiff, L. R. (1998). Cooperative Knowledge Work and Practices of Trust: Sharing Environmental Planning Data Sets. CSCW '98: The ACM Conference On Computer Supported Cooperative Work. Seattle, WA, November 14-18, 1998, Proceedings; ACM, 1998, 335-343. <https://doi.org/10.1145/289444.289508>
- Van House, N. A. (2003). Digital Libraries and Collaborative Knowledge construction. I B. P. Battenfield, A. Peterson-Kemp, & N. A. Van House (Red.), *Digital Library Use: Social Practice in Design and Evaluation* (s. 272-295). Cambridge, MA: MIT Press.
- Van House, N. A. (2004). Science and Technology Studies and Information Studies. *Annual Review of Information Science and Technology*, 38 (1), 3-86.

- Wallis, J. C., Rolando E., Borgman C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE* 8(7), 1-17.
<https://doi.org/10.1371/journal.pone.0067332>
- Walters, W. H. (2020). Data journals: incentivizing data access and documentation within the scholarly communication system", *Insights*, 2020, 33 (18), 1–20.
DOI: <https://doi.org/10.1629/uksg.510>
- Whyte, A. & Pryor, G. (2011). Open science in practice: researcher perspectives and participation. *International journal of digital curation*, 1(6), 199-213.
<https://doi.org/10.2218/ijdc.v6i1.182>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., Santos, L. B. D., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship, *Scientific data*, (3), 1-9, Artikkel 160018.
<https://doi.org/10.1038/sdata.2016.18>
- Yoon, A. (2017). Data Reusers' Trust Development. *Journal of the Association for Information Science and Technology*, 68:4, 946-956. doi: 10.1002/asi.23730
- Zimmerman, A. S. (2003). *Data sharing and secondary use of scientific data: experiences of ecologists*. Unpublished dissertation, University of Michigan, Ann Arbor, MI. Hentet fra [https://www.researchgate.net/publication/30862490 Data Sharing and Secondary Use of Scientific Data Experiences of Ecologists](https://www.researchgate.net/publication/30862490_Data_Sharing_and_Secondary_Use_of_Scientific_Data_Experiences_of_Ecologists)
- Zimmerman A. S., (2008). New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology, & Human Values*, 33(5), 631-652.
DOI:[10.1177/0162243907306704](https://doi.org/10.1177/0162243907306704)

Vedlegg 1.

Vil du delta i forskningsprosjektet

«Deling av forskningsdata – forskerens perspektiv»?

Dette er et spørsmål til deg om å delta i forskningsprosjektet «Deling av forskningsdata – forskerens perspektiv». I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Bakgrunn og formål

Dette prosjektet gjøres som en del av arbeidet med en masteroppgave i bibliotek- og informasjonsvitenskap ved OsloMet - storbyuniversitetet. Prosjektet har til hensikt å undersøke hvilke tanker og erfaringer forskere har rundt deling av forskningsdata. Jeg er interessert i å få vite mer om hvordan slik datadeling fungerer i praksis, og om det er noen spesifikke faktorer som har innvirkning på forskeres datadeling.

Deling av forskningsdata oppfordres for forskning finansiert blant annet av EUs rammeprogram for forskning og innovasjon, Horisont 2020 og forskning finansiert gjennom Norges forskningsråd. Oppfordring om å dele forskningsdata, innebærer at forskeren må vurdere å gjøre informasjon om digitale forskningsdata tilgjengelig og lagre data over lengre tid, slik at de kan gjenbrukes (European Commission 2018, Norges forskningsråd 2017).

Det mangler et klart bilde over hvordan forskere responderer på disse utfordringene, og de utfordringene og problemene de møter når og hvis de gjør forskningsdata tilgjengelige for andre. Hensikten med denne studien er å skape en større forståelse for forskeres ulike utgangspunkt, motivasjoner og opplevde hindringer for å dele data. Ved å intervju forskere om konkrete utfordringer de har møtt og erfaringer de har gjort seg, håper jeg å bidra til mer innsikt i kompleksiteten rundt deling av forskningsdata.

Hvem er ansvarlig for forskningsprosjektet?

Fakultet for samfunnsvitenskap (SAM) ved OsloMet er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Som case har jeg valgt å ta utgangspunkt i polarforskning, siden det både fra forskere selv og fra Forskningsrådets side er nevnt at deling av forskningsdata er spesielt viktig innenfor dette feltet. Valget av informanter til å besvare disse spørsmålene gjøres ved å velge ut forskere som enten har delt forskningsdata i arkiv, eller som har deltatt i forskningsprosjekter innenfor polarforskning.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at du samtykker til at jeg gjennomfører et en-til-en-intervju med deg via Zoom, for å høre om dine erfaringer og tanker rundt det å dele forskningsdata. Intervjuet vil også inneholde noen spørsmål rundt hvilke forskningsmetoder du bruker og hvilke typer data som genereres i tilknytning til forskningen din.

Intervjuet vil ta ca. 60 - 75 minutter å gjennomføre. Intervjuene vil tas opp med lydopptaker, og i tillegg vil det tas notater. Dette for å kunne gjengi informasjonen mest mulig riktig, og for å unngå tap av informasjon. Alle lydopptak vil bli slettet så snart de er gjennomgått og informasjonen er bearbeidet.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

Hva skjer med informasjonen om deg?

Informantens navn vil ikke oppgis i oppgaven, heller ikke hvilken institusjon informanten arbeider ved. Andre opplysninger som intervjuer for rede på under intervjuet som kan identifisere en person vil også bli anonymisert. Derimot kan det komme frem hvilken stilling informanten har, og hvilken fagdisiplin informanten tilhører. Alle personopplysninger vil bli behandlet konfidensielt og oppbevares adskilt fra det innsamlede datamaterialet. Personopplysninger vil kun bli kjent for den som utfører intervjuene, samt prosjektets veileder Nils Pharo.

Intervjumaterialet som samles inn i studien vil aidentifiseres og informantens navn vil erstattes med koder. Alle lydopptak slettes når oppgaven er ferdigstilt. Prosjektet skal etter planen avsluttes desember 2021.

Hva gir oss rett til å behandle personopplysninger om deg?


Vi behandler opplysninger om deg basert på ditt samtykke. På oppdrag fra OsloMet har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

- Masterstudent: Karen Lillo. Kan kontaktes på e-post: s149069@oslomet.no eller mobil: 
- Veileder for oppgaven: professor Nils Pharo på e-post: nils.pharo@oslomet.no
- Vårt personvernombud: Ingrid Jacobsen på e-post: personvernombud@oslomet.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen
Nils Pharo
Prosjektansvarlig
(veileder)

Karen Lillo
student

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «*Deling av forskningsdata – forskerens perspektiv*», og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å delta i intervju

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

(Signert av prosjektdeltaker, dato)

Vedlegg 2.

Intervjuguide

Introduksjon

Takk for at du vil være med å bidra i dette prosjektet.

Kort introduksjon av meg og forskningsprosjektet.

Informasjon om fortrolighet og personvern

Informere om lydopptak og at samtykke kan trekkes tilbake.

Bakgrunnsspørsmål

1. Hva er din stilling?
2. Hvor lenge har du vært involvert i forskning?
3. Hvilken fagdisiplin tilhører du?

Forskningsfelt

1. Kan du kort beskrive dine hovedområder innen forskning? Hva er de beste begrepene/nøkkelordene for å beskrive dette forskningsområdet?
2. Jeg prøver å finne ut omfanget av fagdisiplinen din? Er det mange som jobber innen dette feltet i Norge? Eller er det bare noen få eksperter? Hva med på et internasjonalt nivå?
3. I hvilken grad er det enighet i feltet om de viktigste forskningsspørsmålene, tilnærmingene og metodene?
4. Låner feltet begreper/metoder fra andre fagfelt? (Hvor tverrfaglig er feltet?)
5. Forsker du mest alene eller i samarbeid med andre?
6. Gjør du ofte forskning med andre forskere innen samme fagområde eller utover eget fagområde, eller begge deler? [innen avdelingen / Norge / internasjonalt?] Hvorfor? (Er det vanlig praksis?) (Hva slags disipliner kommer de fra?)

7. Hvordan vil du beskrive forskningsgruppens betydning for deg? (Hvordan påvirker det forskningen din?)

Jeg vil nå gå videre til å diskutere datapraksis og dataenes karakter.

1. Kan du kort forklare hvilke forskningsmetoder du vanligvis bruker? Hvilke typer data genererer/skaper du i tilknytningen til forskningen din? (Primærdata) (Kvantitative- hva slags?), (Kvalitative – hva slags?). (Eller begge deler?).
2. Hvordan behandler du dataene dine (f.eks. hvor får du dataene dine fra, hvilke verktøy brukes til å analysere eller behandle dem?)
3. Hva skjer med dataene vanligvis etter publisering av tidsskriftartikler eller gjennomføring av et forskningsprosjekt?
 - a) Blir dataene lagret? Hvor? Hvor lenge? Tar du vare på dem for videre forskning? Bruker du egne data i andre forskningsprosjekter?
 - b) Blir dataene gjort tilgjengelig for andre?
4. Er det noen forskjeller når det gjelder hva som kreves for å håndtere forskjellige typer data? (Er for eksempel noen typer data vanskeligere å arkivere eller dokumentere enn andre?)
5. Vet du om noen spesifikke faktorer som virker inn på datadelingspraksis innen ditt forskningsområde?

Datadeling generelt

1. Hvordan deler du vanligvis dataene dine når du samarbeider med andre forskere om et forskningsprosjekt? f.eks. hvilke verktøy bruker du? Hva slags data deler du? Fungerer det bra?
2. Hva synes du om idéen om å dele forskningsdata utenfor forskningsgruppen/teamet ditt?
 - Vilje til å dele data?
 - Konkurransen innen feltet?
 - Verdien/ nødvendighet av å dele data innen ditt felt?
 - Copyright- forfatterskap?
3. Har du noen betenkeligheter rundt det å dele egne forskningsdata (for eksempel etiske årsaker, eller andre typer risiko)?
4. Har du noen gang mottatt forespørsel om å dele dataene dine? (f.eks koding, transkripsjoner, bilder etc.)

5. Hvilke typer data har du delt (rådata)?
6. Deler du alle dine data, eller finnes det data du har valgt å ikke dele, og i så fall hvorfor? (Hvilke data deles ikke?)
7. Når er det aktuelt å dele data? (Når - tidsmessig –etter publisering, gjør du en vurdering av nytteverdien for andre?)
8. Under hvilke forhold eller betingelser er du villig til å dele egne data? (Krav fra finansører, forskere fra samme forskningsfelt, etter publisering av publikasjoner). (Ønsker du eller er det behov for å kontrollere hvem som får tilgang til dataene dine?).
9. Hva ville motivert deg til å dele dine forskningsdata? (f.eks policyer, datasitering, fagspesifikke arkiv, økt synlighet..)

Deling av data i arkiv/databaser

1. Har du eller forskningsgruppen din delt data i arkiv? Hva var motivasjonen for å dele data i arkiv?
2. Hva synes du er de største barrierene eller hindringene for å gjøre data tilgjengelig i arkiv?
3. **Hvis deling av data i arkiv:** Hvordan gikk det til rent teknisk når du skulle dele data i arkiv, støtte du på noen vanskeligheter?
4. Har du noen målgruppe i tankene når du deler data i arkiv?
5. Opplever du at andre (forskere) tar kontakt i forhold til datasettene du har gjort tilgjengelig i arkiv? (Forskere innen samme fagområde/disiplin, forskere fra andre fagområder, andre?) Må du forklare noe? (Hva er det vanlig at de spør om?)
6. Oppstår det noe interaksjon? Forespørsel om samarbeid?
7. Har du noe inntrykk av om dataene som er gjort tilgjengelig i arkiv av deg eller din forskningsgruppe, blir gjenbrukt?
8. Kunne du tenke deg å dele flere data i arkiv, som er ment for å kunne gjenbrukes av andre?

Sitering av datasett (hvis deling av data i arkiv)

1. Har du opplevd at datasett du har delt, har blitt sitert?

2. Er det vanlig praksis innen ditt felt å sitere datasett?
3. Er det noen forskjell på å sitere datasett, jamfør det å sitere en vanlig tradisjonell vitenskapelig artikkel?

Mekanismer rundt deling av forskningsdata

1. Hva tror du er de største hindringene for datadeling blant forskerne?
2. Finnes det noe som skulle kunne gjøre det lettere å dele data?
3. Er det gode nok insentiver- og belønningssystemer for å dele data?
4. Har noen policy (nasjonal, finansørers, egen institusjon, forlag/tidsskrift) hatt betydning for at du har delt data?
5. Hva slags kultur er det rundt deling av forskningsdata innen ditt forskningsområde?

Datagjenbruk

Et annet viktig tema jeg ønsker dine synspunkter på er gjenbruk av forskningsdata (bruk av andres forskeres data).

1. Har du opplevd å ha behov for data som er generert av andre forskere som ikke er tilgjengelig?
2. Bruker du data generert/samlet inn av andre forskere? Hva slags data er dette? Hvordan bruker du (eller evt. hvorfor ikke) disse dataene på nytt? (Er det vanlig praksis innen ditt felt?). (Nasjonale/internasjonale datasett. Er de fritt tilgjengelige?)
3. Gjør du en vurdering kvaliteten på dataene? Hva er dine kriterier for å kunne gjenbruke data? (f. eks forskningsfelt/integritet)
4. Er det noen utfordringer med å gjenbruke forskningsdata fra andre?

Trender

I litteraturen er det beskrevet noen trender som jeg lurer på om du merker noe til innenfor din forskning:

Krav om eller oppfordring om tverrfaglighet/interdisiplinæritet.

Teknologi – økende grad av store, komplekse datasett, - såkalte big data.

Vedlegg 3.



Zoom i forskningsintervjuer – Informasjonsskriv

Du er invitert som deltager i et forskningsintervju som skal gjennomføres i videoverktøyet Zoom. I dette skrevet gir vi deg informasjon om hva deltakelsen vil innebære for deg. Det følger også med praktisk informasjon om gjennomføring av intervjuet.

Du må ha lest denne informasjonen før du samtykker til å delta i intervjuet og før du samtykker til et eventuelt opptak av intervjuet.

Om intervjuet

Møtet gjennomføres som et enkeltintervju.

Formål

Dette er et spørsmål til deg om å delta i forskningsprosjektet «Deling av forskningsdata – forskerens perspektiv». Dette prosjektet gjøres som en del av arbeidet med en masteroppgave i bibliotek- og informasjonsvitenskap ved [OsloMet](#) storbyuniversitetet. Prosjektet har til hensikt å undersøke hvilke tanker og erfaringer forskere har rundt deling av forskningsdata. Jeg er interessert i å få vite mer om hvordan slik datadeling fungerer i praksis, og om det er noen spesifikke faktorer som har innvirkning på forskeres datadeling og datapraksis.

Deling av forskningsdata oppfordres for forskning finansiert blant annet av EUs rammeprogram for forskning og innovasjon, Horisont 2020 og forskning finansiert gjennom Norges forskningsråd. Oppfordring om å dele forskningsdata, innebærer at forskeren må vurdere å gjøre informasjon om digitale forskningsdata tilgjengelig og lagre data over lengre tid, slik at de kan gjenbrukes (European Commission 2018, Norges forskningsråd 2017).

Det mangler et klart bilde over hvordan forskere responderer på disse utfordringene, og de utfordringene og problemene de møter når og hvis de gjør forskningsdata tilgjengelige for andre. Hensikten med denne studien er å skape en større forståelse for forskeres ulike utgangspunkt, motivasjoner og opplevde hindringer for å dele data. Ved å intervjuere forskere om konkrete utfordringer de har møtt og erfaringer de har gjort seg, håper jeg å bidra til mer innsikt i kompleksiteten rundt deling av forskningsdata.

Hvorfor får du spørsmål om å delta?

Som case har jeg valgt å ta utgangspunkt i polarforskning, siden det både fra forskere selv og fra Forskningsrådets side er nevnt at deling av forskningsdata er spesielt viktig innenfor dette feltet. Valget av informanter til å besvare disse spørsmålene gjøres ved å velge ut forskere som enten har delt forskningsdata i arkiv, eller som har deltatt i forskningsprosjekter innenfor polarforskning

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at du samtykker til at jeg gjennomfører et en-til-en-intervju med deg via Zoom, for å høre om dine erfaringer og tanker rundt det å dele forskningsdata. Intervjuet vil også inneholde noen spørsmål rundt hvilke forskningsmetoder du bruker og hvilke typer data som genereres i tilknytning til forskningen din.

Intervjuet vil ta ca. 60 - 75 minutter å gjennomføre. Intervjuene vil tas opp med lydopptaker, og i tillegg vil det tas notater. Dette for å kunne gjengi informasjonen mest mulig riktig, og for å unngå tap av informasjon. Alle lydopptak vil bli slettet så snart de er gjennomgått og informasjonen er bearbeidet.

Hva skjer med informasjonen om deg?

Informantens navn vil ikke oppgis i oppgaven, heller ikke hvilken institusjon informanten arbeider ved. Andre opplysninger som intervjuer for rede på under intervjuet som kan identifisere en person vil også bli anonymisert. Derimot kan det komme frem hvilken stilling informanten har, og hvilken fagdisiplin informanten tilhører. Alle personopplysninger vil bli behandlet konfidensielt og oppbevares adskilt fra det innsamlede datamaterialet. Personopplysninger vil kun bli kjent for den som utfører intervjuene, samt prosjektets veileder Nils Pharo.

Intervjumaterialet som samles inn i studien vil aidentifiseres og informantens navn vil erstattes med koder. Alle lydopptak slettes når oppgaven er ferdigstilt. Prosjektet skal etter planen avsluttes desember 2021.

Opptak og notater fra intervjuet

- Det vil bli tatt opptak av lyd i annet verktøy.
- Det blir ikke gjort opptak av intervjuet, bare notater.

Dersom det gjøres opptak, vil alt du sier bli med på opptaket.

Det er frivillig å delta

Det er frivillig å delta i intervjuet. Hvis du velger å delta/svare, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn.

Det er ikke mulig å trekke tilbake samtykket etter at intervjuet er gjennomført, men du kan be om at opptaket eller notatene av det du har sagt slettes.

Alle dine personopplysninger og utsagn vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke i henhold til personvernforordningen artikkel 6 og artikkel 9.

- ved alminnelige personopplysninger artikkel 6 a) «den registrerte har samtykket til behandling av sine personopplysninger for ett eller flere spesifikke formål»,
- ved særlige kategorier som f.eks. helseopplysninger, fagforening eller liknende artikkel 6 a) og artikkel 9, nr 2 a) «Den registrerte har gitt uttrykkelig samtykke til behandling av slike personopplysninger for ett eller flere spesifikke formål, unntatt dersom det i unionsretten eller medlemsstatenes nasjonale rett er fastsatt at den registrerte ikke kan oppheve forbudet nevnt i nr. 1.»

Ansvarlig

OsloMet er behandlingsansvarlig for innsamlingen og for behandlingen av dine personopplysninger.

Kontaktpersoner

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

- Masterstudent: Karen Lillo. Kan kontaktes på e-post: s149069@oslomet.no eller mobil: [REDACTED]
- Veileder for oppgaven: professor Nils Pharo på e-post: nils.pharo@oslomet.no
- Vårt personvernombud: Ingrid Jacobsen på e-post: personvernombud@oslomet.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Nils Pharo
Prosjektansvarlig
(veileder)

Karen Lillo
student

Praktisk informasjon om gjennomføring av møtet

Zoom-appen

Dersom du ikke allerede har lastet ned Zoom-appen, anbefaler vi at du gjør det før møtet starter. Du finner den her:

PC/MAC: <https://oslomet.zoom.us/download>

Telefon/Nettbrett: <https://apps.apple.com/us/app/id546505307>
<https://play.google.com/store/apps/details?id=us.zoom.videomeetings>

Det er også mulig å delta via nettleser. Chrome og Safari er anbefalt. Det vil ikke fungere i Edge og Internet Explorer.

Forberedelser

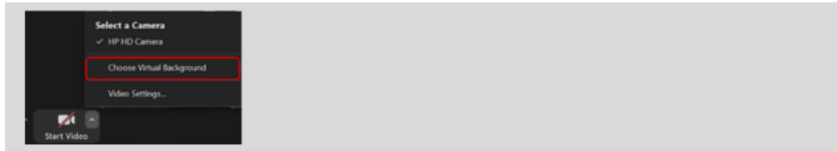
Pass på at du sitter på et sted hvor ingen kan høre lyd fra samtalen eller se skjermen din. Bruk hodetelefoner om du har.

Skru av video og endre visningsnavn

Dersom du ønsker, kan du skru av video i det du går inn i møtet. Klikk på «Join without Video».

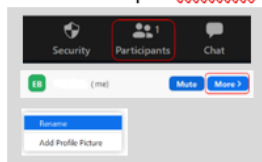


Om du ønsker å delta med video, kan bruke en virtuell bakgrunn for å skjule omgivelsene du sitter i. Åpne videomenyen og klikk på «Choose Virtual Background»:



Du kan også endre ditt eget visningsnavn.

- Klikk på «Participants».
- Før musepekeren over ditt eget navn og klikk «More».
- Klikk på «Rename» og endre ditt visningsnavn.



Dele skjerm

Dersom du skal dele skjerm, bør du lukke applikasjoner og filer du ikke vil vise.

Optak

Du vil få varsel i zoom dersom det gjøres opptak. NB! Av både personvern hensyn og opphavsrettslige hensyn skal du ikke selv ta opptak av intervjuet.

Samtykkeskjema – Samtykke til forskningsintervju i Zoom

Før du samtykker må du ha lest informasjonsskrivet *Zoom i forskningsintervjuer- Informasjonsskriv*, slik at du er orientert om formålet med intervjuet og rettighetene dine.

Jeg samtykker til å delta i forskningsintervju på zoom i forbindelse med forskningsprosjektet «Deling av forskningsdata – forskerens perspektiv».

Jeg samtykker til at det vil bli tatt opptak av lyd i Nettskjema-diktafon-appen.

Dato: Sted:

Navn (Blokkbokstaver) _____

Navn (Underskrift) _____

Vedlegg 4.

Melding

10.05.2021 10:07

Behandlingen av personopplysninger er vurdert av NSD. Vurderingen er:

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 10.05.2021. Behandlingen kan starte.

DEL PROSJEKTET MED PROSJEKTANSVARLIG

For studenter er det obligatorisk å dele prosjektet med prosjektansvarlig (veileder). Del ved å trykke på knappen «Del prosjekt» i menylinjen øverst i meldeskjemaet. Prosjektansvarlig bes akseptere invitasjonen innen en uke. Om invitasjonen utløper, må han/hun inviteres på nytt.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 20.12.2021.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål
dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), og dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

Zoom og Nettskjema er databehandlere i prosjektet. NSD legger til grunn at behandlingen oppfyller kravene til bruk av databehandler, jf. art 28 og 29.

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

MELD VESENTLIGE ENDRINGER

Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: nsd.no/personverntjenester/fyle-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema
Du må vente på svar fra NSD før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET

NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!