

Kandidatnummer 103

Forskningsdata på Handelshøyskolen BI
Hva er forskernes behov?

Masteroppgave 2015
Master i bibliotek, styring og ledelse
Høgskolen i Oslo og Akershus, Institutt for arkiv-, bibliotek- og informasjonsfag

Forskningsdata på Handelshøyskolen BI – Hvilket behov har forskerne?

BSL5900 Masteroppgave

Innhold

Abstract	iv
Sammendrag	v
Forkortelser / organisasjoner	vi
Forord	vii
Introduksjon	9
Problemstilling	11
Hoveddel	12
Hva er forskningsdata?	12
Forskningspolitiske føringer og retningslinjer for arbeid med forskningsdata	15
Data, informasjon og kunnskap	19
Vitenskapens verdier	24
Forskeres syn på lagring av forskningsdata	29
Forskningsdata i praksisfeltet	35
Utfordringer for forskningsinstitusjonene	38
Data Asset Framework	39
Kort beskrivelse av Handelshøyskolen BI	39
Metode	41
Valg av metode	41
Utvalget	41
Det kvantitative undersøkelsesdesignet; spørreundersøkelsen	42
Beskrivelse av spørreundersøkelsen	43
Pilottest av undersøkelsen	44
Kvalitet i undersøkelsen	45
Generalisering	46
Etiske betraktninger	46

Gjennomføring av undersøkelsen	47
Resultater	49
Diskusjon	70
Endringer i spørreskjema	76
Anbefalinger	79
Infrastruktur	80
Mer ressurser	82
Opphavsrett og lisenser.....	83
Opplæring.....	84
Arbeidsgruppe	85
Funn, implikasjoner og tiltak fra undersøkelsen	86
Konklusjon	89
Appendiks.....	92
Appendiks A. Foreslåtte tiltak fra Norges forskningsråd (2014 a).	92
Appendiks B. Invitasjonsbrev	93
Appendiks C. Utsendt spørresundersøkelse	94
Appendiks D. Årsaker til bevaring av data.....	100
Appendiks E. Kommentarer om tema forskningsdata eller undersøkelsen.....	102
Appendiks F. Forskningsdatas gang i forskningsprosessen	103
Referanser	106

Figurliste

Figur 1. Filformater og reproduksjon ved tap av data.....	55
Figur 2. Lagringsmåte for data sett i forhold til mulighet for å gjenskape data dersom data i en undersøkelse forsvinner	58
Figur 3. Bruk av lagringsenheter fordelt på avdelingstilhørighet.....	59
Figur 4. Problems, solutions and recommendations «Sharing Research Data» (Norges forskningsråd, 2014a).....	92
Figur 5. Forskningsdata ved prosjektetablering	103
Figur 6. Forskningsdata ved prosjektgjennomføring.....	104
Figur 7. Forskningsdata ved prosjektavslutning	105

Tabelloversikt

Tabell 1. Respons basert på avdelingsvis fordeling	49
Tabell 2. Respondentenes stillingskategori og avdelingstilknytning	50
Tabell 3. Finansieringskilder for pågående prosjekter blant respondentene	51
Tabell 4. Datatyper basert på innsamlingsmetode	52
Tabell 5. Filformater som genereres av forskerne	53
Tabell 6. Estimerte filstørrelse for respondentenes forskningsprosjekter	54
Tabell 7. Vanskeligheter ved å gjenskape data	54
Tabell 8. Forskeres bruk av data produsert av andre forskere	56
Tabell 9. Dataarkivering etter avsluttet prosjekt benyttet av forskerne	57
Tabell 10. Manglende tilfredshet med dagens arkiveringsløsning	60
Tabell 11. Deling av anonymiserte data	61
Tabell 12. Utfordringer med å gjøre data tilgjengelig for andre forskere	62
Tabell 13. Forskernes syn på data som del av vitenskapelige publikasjoner	63
Tabell 14. Respondentenes beregning av hvor lenge data bør lagres	63
Tabell 15. Hyppige argumenter for behov for lagring og bevaring av data	65
Tabell 16. Forskernes ønsker om tiltak ved BI	66
Tabell 17. Ønske om tiltak for BI fordelt over forskernes avdelingstilhørighet	67
Tabell 18. Forskernes ønsker om tiltak for egen utvikling	68
Tabell 19. Hovedfunn med forslag til anbefalinger og tiltak	87

Abstract

Research Data Management at BI Norwegian Business School – Researchers’ needs

Researchers’ data practices were assessed to find how the higher education research institution BI Norwegian Business School can cater for researchers’ needs.

59 researchers answered an online survey about their current research data practice, thoughts on data curation and thoughts on how their research institution should enhance work processes for research data. The survey was tailored to a Norwegian context based on Parade Open Access to Research Data Survey (L. H. Kvale, 2012; Norges forskningsråd, 2014a) and the Data Asset Framework (Jones, Ross, & Ruusalepp, 2009; Rolando, Doty, Hagenmaier, Valk, & Parham, 2013).

Researchers have a positive attitude towards sharing data. Obstacle toward data sharing are sensitive data, time to prepare data for open access and fear for missed publishing opportunities. Researchers would like to share data with fellow researchers, but to a limited extent to the general public.

Most researchers are not part of the data intensive e-science, using theoretical and experimental research designs producing small file sizes (>500 GB) in numerical or text formats. After project termination research data are saved on personal storage solutions or institutional servers. Data archives are used to a limited extent. Researchers are content with their current research data management practice.

Proposed changes are related to data storage and exchange in ongoing projects and better infrastructure. The researchers are interested in training and information on best practice research data management.

Keywords

Researchers, Research Data Management, Norway, Research Institutions, Higher Education

Sammendrag

Forskningsdata på Handelshøyskolen BI – Hvilket behov har forskerne?

Forskeres data praksis var kartlagt for å finne hvordan den høyere utdanningsinstitusjonen Handelshøyskolen BI kan ivareta forskernes behov.

59 forskere besvarte et elektronisk spørreskjema om deres datapraksis, tanker om databevaring og hvordan deres forskningsinstitusjon skal forbedre arbeidsprosessene for forskningsdata. Undersøkelsen var tilpasset en norsk kontekst ved å anvende Parade Open Access to Research Data Survey (L. H. Kvale, 2012; Norges forskningsråd, 2014a) og Data Asset Framework (Jones et al., 2009; Rolando et al., 2013).

Forskere har en positiv holdning til å dele data. Hindre for deling av data er forekomst av sensitive data, tid for å gjøre data åpent tilgjengelig og reduksjon av muligheten for fremtidig publisering. Forskere vil gjerne dele data med andre forskere, men i liten grad med allmennheten.

De fleste forskere er ikke en del av en data intensiv e-science. De anvender teoretiske og eksperimentelle forskningsdesign som produserer små filstørrelser (>500 GB) i formater som regneark og tekstdokumenter. Etter prosjektavslutning lagres data lokalt på personlige lagringsmedium eller institusjonelle servere. Dataarkiver er i liten grad brukt. Forskeren er fornøyd med dagens lagringsløsninger.

Foreslåtte endringer er knyttet til datalagring og utveksling av data for pågående forskningsprosjekter og tilgang til bedre infrastruktur. Forskerne er interessert i opplæring og diskusjoner om beste praksis for administrasjon av forskningsdata.

Nøkkelord

Forskere, forskningsdata, Norge, forskningsinstitusjoner, høyere utdanning

Forkortelser / organisasjoner

FORKORTEELSE	ORGANISASJON/BESKRIVELSE
BI	Handelshøyskolen BI
CARDIO	Collaborative Assessment of Research Data Infrastructures and Objectives
CRISTIN	Current Research Information System in Norway
DAF	Data Asset Framework (tidl. Data Audit Framework)
DCC	Digital Curation Center
DMP	Data Management Plan
LIBER	US Council on Library and Information Resources
EUDAT	European Data Infrastructure
EDAWAX	European Data Watch Extended
HATII	Humanities Advanced Technology & Information Institute
HEI	Higher Education Institutions
JISC	Joint information Systems Collaboration
LIS	Library and Information Science
LIBER	The Association of European Research Libraries
NEIC	Nordic eInfrastructure Collaboration
NSD	Norsk Samfunnsvitenskapelige Datatjeneste
OA	Open Access
OECD	Organisation for Economic Co-operation and Development
PARADE	Partnership for Advanced Data in Europe
PARSE	Permanent Access to the Records of Science in Europe
RDM	Research Data Management
REF	Research Excellence Framework
RIN	Research Information Network
RDA	Research Data Alliance
RLUK	Research Libraries United Kingdom
SSB	Statistisk sentralbyrå

Forord

Første gang jeg hørte at forskningsdata kan være en oppgave for bibliotek var i 2009 på et seminar ved Universitetet i Bergen. Stemningen for forslaget var kjølig fra bibliotekarene i salen på Studentsenteret. De fleste visste lite om hva arbeid med forskningsdata innebar. Nå, 6 år senere, er det interessant å prøve å finne ut hva bibliotekene mener om forskningsdata. Er dette en oppgave som bibliotekene skal involvere seg i? Tilbyr noen bibliotek tjenester i tilknytning til forskningsdata? Planlegger bibliotek å starte med denne typen tjenester?

I løpet av dette året har min læringskurve omkring forskningsdata vært sterkt stigende. Internasjonalt har flere bibliotekorganisasjoner pekt på at forskningsdata er en oppgave som kommer sterkere for bibliotekene i løpet av de kommende årene. Området forskningsdata berører de fleste felt innen bibliotek; samlingsutvikling, undervisning- og forskningsstøtte, teknologiske og digitale tjenester, strategi og ledelse.

Ved oppstarten av arbeid med forskningsdata reises der flere spørsmål: Hva gjøres i dag? Hvordan bør dette organiseres? Hvilke roller skal ulike aktører ha? Hvilke aktiviteter skal vi beskjeftige oss med? Hva er beste praksis?

Forskningsdata kan studeres fra flere innfallsvinkler: politiske føringer fra myndigheter, forskningsfinansiørers ønsker, økonomiske betingelser og behov, teknologiske systemer, høyere utdanningsorganisasjoner eller forskningsinstitutters behov, forsker- og brukerperspektiv. I tillegg kan man se på ulike aktører som opererer innen de nevnte feltene eller fokusere på aktiviteter innen ulike faser av lagring, forvaltning og deling av forskningsdata.

Fagmiljøet ved bibliotek- og informasjonsvitenskap diskuterer om forskningsdata er en egen profesjon. Det store behovet for spesialisert kompetanse og ferdigheter har utløst diskusjonen. Er arbeidet med forskningsdata en del av bibliotek- og informasjonsvitenskapen hvor man gjennom en spesialisering kvalifiserer til å arbeide med forskningsdata? Eller er forskningsdata et arbeid for forskere som tar en spesialisering?

Forskningsdata er et område som får oppmerksomhet i dag, og det kommer til å være i fokus for universitets- og høyskolesektoren i årene fremover. Bibliotekene samhandler med fagmiljø

og ledelse i sektoren for å nå organisasjonenes felles mål for forskning og undervisning. Bibliotekene har i lengre tid hatt fokus på å skreddersy sitt tilbud til sine brukergruppers behov. Forskernes behov i sitt arbeid med forskningsdata er interessant. Et kunnskapsbasert grunnlag vil bidra til å kunne legge grunnlag for strategiske ledelsesbeslutninger og fremtidige tjenester. Hvordan bibliotekene eventuelt velger å ta initiativ til en rolle på feltet gjenstår å se, uansett vil jeg tro forskningsdata blir et tema i forbindelse med bibliotekenes forskningsstøttetjenester på sikt.

Interesserte kan se nærmere på internasjonal biblioteksektors tanker om forskningsdata som en bibliotekjeneste for fremtiden (Aukland, 2012; Laursen, Brøndum, Davidsen, Hagelskjær, & Svensson, 2013; Nielsen, 2012), hvilke roller dette potensielt kan innebære, (Reilly, Schaller, Schrimpf, Smit, & Wilkinson, 2011; Thestrup et al., 2012; Van der Graaf & Waaijers, 2011), bibliotekenes nåtidige tjenester og planer (Cox & Pinfield, 2013; C. Tenopir, Sandusky, Allard, & Birch, 2014).

Introduksjon

Forskere i Norge møter nye utfordringer med henhold til datalagring og deling gjennom forskningspolitiske krav fremsatt til forskningsinstitusjoner, krav fra forskningsfinansiører omkring tilgjengeliggjøring av data i arkiv, eller krav fra forlag om publisering av data ved vitenskapelig publisering av resultater.

Universitets- og høyskolesektoren opplever at politiske myndigheter har signalisert ønske om større grad av åpenhet og krav til deling av forskningsdata. I Forskningsmeldingen kap. 3.4.3. sies det at “Bedre tilgang til forskningsdata bidrar til å forenkle forskningssamarbeid og til å høyne kvaliteten på forskningen. Regjeringen ønsker å legge til rette for økt tilgjengeliggjøring av offentlige forskningsdata” (Kunnskapsdepartementet, 2013). Dette er i henhold til de signaler som er gitt internasjonalt av Organisation for Economic Co-operation and Development (2007) [OECD].

Finansieringskilden Norges forskningsråd [NFR] lanserte i september 2014 retningslinjer for arkivering og deling av forskningsdata. Hovedregelen er at forskningsdata innhentet over midler fra NFR skal stilles tilgjengelig for andre forskere (Norges forskningsråd, 2014b). Dette er en forlenging av arbeidet med Norges Forskningsråds nasjonale strategi for forskningsinfrastruktur og et nytt veikart over strategisk viktige infrastrukturer (2012). Lignende krav om tilgjengeliggjøring av forskningsdata stilles ved søknader til European Commission (2013) [EU] gjennom Horizon 2020.

I forskningssektoren omtales prosessen med produksjon, bearbeiding og analyse av data, lagring og bevaring som «Research Data Management» [RDM]. Research Data Management er en konsekvens av teknologiske endringer og utvikling i retning av tverrfaglig forskningssamarbeid på globalt nivå, samt policy som følger myndighetenes behov for etterprøvbarhet av deres investeringer i forskning (Corrall, Kennan, & Afzal, 2013).

For den enkelte forskningsinstitusjon vil klargjøring omkring hvilke roller og aktiviteter som de ulike aktørene kan bidra med være viktig for å vite hvordan de kan bistå forskerens behov for forskningsinfrastruktur. Internt ved forskningsinstitusjonene berører arbeidet ledelsen og forskningsadministrasjon, forskere og forskningsgrupper, IT- avdeling og eventuelt bibliotek. Utenfor forskningsinstitusjon involverer arbeidet private eller offentlige tjenesteleverandører,

arkiver for datalagring, myndigheter og forskningsfinansiører (f.eks. Michener et al 2012, ref i Whyte, 2014, s.73).

National Science Board (2005) har definert de ulike aktørene med roller ved digitale data. Aktørene omtales som databrukere, dataprodusenter, dataadministratorer og finansieringsinstanser. Dataprodusenter, “data authors”, er forskere, studenter eller andre som er involvert i forskningen som produserer digitale data. Dataadministratorer, “Data managers”, er personer eller organisasjoner ansvarlig for opprettelse og vedlikehold av databaser. “Data scientists” er personer som gjør analyser og sørger for å ivareta digitale data samlinger. Dette kan være spesialister på programvare, datamaskiner, arkivarer, bibliotekarer eller andre. Databrukerne, “data users”, er en henvisning til det øvrige forsknings-samfunnet som anvender data som er samlet inn. Rollene er ikke skarpt definerte, slik at en forsker som er dataprodusent, kan i tillegg være en databruker ved at forskeren gjenbraker data som andre har samlet inn. Når forskere omtales i denne sammenheng, vil det først og fremst være en henvisning til rollen som dataprodusent. Brukere vil henviser til andre personer som gjenbraker data. IT-avdelinger og arkiver vil i de fleste tilfelle ha rolle som dataadministratorer.

Lagring og tilgjengeliggjøring av data krever spesialiserte kunnskaper og ferdigheter, og arbeid med forskningsdata vil være kostbart for utdanningsinstitusjonen. Norges forskningsråd (2014a) fremstiller flere mulige tiltak for å få forskere til å avlevere og dele data. Ett av de nevnte tiltakene er bedre informasjon om hva som trengs for at infrastrukturen skal bli bedre. Analyser for å avklare behov er viktig, da investeringer i infrastruktur er kostbart. Kartlegging fremheves som viktig av den europeiske universitetssektoren (League of European Research Universities, 2013) og arkivsektoren som arbeider inn mot forskningsinstitusjoner (Jones, Pryor, & Whyte, 2013).

Per dags dato har vi begrenset kunnskap om datapraksis høyere forsknings- og utdanningsinstitusjoner i Norge. Undersøkelsene som er utført til nå fokuserer på datapraksis i tilknytning til deling av forskningsdata (f.eks L. H. Kvale, 2012; Norges forskningsråd, 2014a). Mangel på teknisk infrastruktur, tid til å forbedre data for arkivering og frykt for at open access til forskningsdata skal hindre vitenskapelig publisering i fremtiden er avgjørende faktorer (Norges forskningsråd, 2014a). I tillegg fremheves det at forskere på forskjellige fagområder kan ha ulike behov på grunn av forskningsmetoder, standarder og dataformater som benyttes av forskerne (Thestrup et al., 2012).

Det finnes en rekke løsninger for lagring av forskningsdata i Norge og internasjonalt som forskere ved høyere utdanningsinstitusjoner benytter seg av. Denne undersøkelsen vil fokusere på den enkelte forskerens aktiviteter innen forskningsdata, med spesiell vekt på rollen fra produsert til avlevering av data for arkivering i tilknytning til en forskningsinstitusjon. Det legges vekt å kartlegge datapraksis og få innspill til forbedrende praksis/tiltak i organisasjonen. For forskningsinstitusjonene er dette aktuelt for å kunne utøve praksis i tråd med myndigheter og forskningsfinansiørers krav og samtidig ivareta forskernes interesser. Andre områder for hvordan universitets- og høyskolesektoren kontrollerer og administrerer forskningsdata faller utenfor undersøkelsen. I midlertid kan resultatene være med på å gi andre aktører et innblikk i utfordringer for forskerne som dataprodusenter møter. Mulige implikasjoner av resultatene kan være med på å danne et utgangspunkt for samarbeid.

Formålet med studien er å kartlegge forskningsdata som finnes ved Handelshøyskolen BI, hvor mye av forskningen er knyttet til fagområdene finans, økonomi og arbeids- og organisasjonsfag.

Målet er å finne ut hvilke dataformater og -typer som finnes, hvordan data lagres og eventuelt deles og gjenbrukes. Dette kan bidra til å identifisere forslag til hvordan man kan forbedre eksisterende praksis i organisasjonen.

Problemstilling

- Hvilke forskningsdata finnes på Handelshøyskolen BI?
 - Hvilke filformater og datatyper brukes i organisasjonen?
 - Hvordan blir data lagret?
 - Hvilken praksis finnes det for å dele data?
 - Hvilken praksis finnes det blant forskerne for å gjenbruke andres data?
- Hvordan kan Handelshøyskolen BI som institusjon bidra til forskernes arbeid med data?
 - Hvilke tiltak for forskningsdata mener forskerne BI bør initiere?
 - Hvilke tiltak for forskningsdata mener forskerne er mest aktuelt for seg selv?

I den videre presentasjonen vil aktuelle definisjoner legges frem, da ulike aktører innen forskningsprosessen kan ha forskjellig forståelse av begrepene som anvendes. Utgangspunktet vil være et informasjonsvitenskapelig perspektiv, hvor begrepsmessige definisjoner av data og informasjon presenteres og settes inn i en forskningspolitisk bakgrunn. Konteksten for undersøkelsen er forskere ved Handelshøyskolen BI. De politiske føringene som ligger til

grunn påvirker forskere og forskningsinstitusjoner i stor grad. Arbeid med research data management er knyttet til alle faser i forskningsprosessen, før under og etter et forskningsprosjekt. Det er derfor av nytte å se nærmere på hvordan forskningsdata er knyttet til gangen i forskningsprosessen. Deretter vil det trekkes vekslers på kunnskapssosiologiske perspektiver om vitenskapens verdier for å knytte dette opp mot forskeres praksisutøvelse. Tilslutt trekkes det inn empiriske undersøkelser omkring forskernes syn på lagring og deling av forskningsdata, hvor foreslåtte tiltak kan ha implikasjoner for forskningsinstitusjoner.

Hoveddel

Hva er forskningsdata?

Tilnærmingen til data er forskjellig, basert på det ståsted som interessegruppene har. Forskningsdata har ikke en entydig definisjon. Definisjoner fra myndigheter og forskningsfinansiører, arkivinstanser og forskningsutøvelse vektlegger ulike aspekter ved data. Forskerne tar gjerne utgangspunkt i metodologiske aspekter, myndigheter og forskningsfinansiører i datas opprinnelse og form, mens arkivene vektlegger aktiviteter i datas livssyklus. Dette vil belyses videre her.

Innenfor forskningsutøvelse sees definisjoner av data basert på metodologiske prinsipper og fremgangsmåter. Østbye, Helland, Knapskog, og Larsen (2013) viser til at forskning skal si noe om virkeligheten, hvor materialet som samles inn ofte omtales som data. Data grupperes gjerne i kvalitative data eller kvantitative data. Kvantitative data kan tallfestes eller telles, mens kvalitative data er erfaringsmateriale.

Forskningsmetodisk fremgangsmåte relateres til paradigmer innen vitenskapen, hvor det beskrives at det eksisterer 4 paradigmer i dag. De to første paradigmene er henvisning til eksperimentell og teoretisk vitenskap, som har eksistert siden 1600-tallet. I løpet av de siste tiårene er beregningsvitenskap, basert på datasimuleringer, kommet til som et paradigme. Nå er det i ferd med å utvikle seg et fjerde paradigme som dataintensive vitenskaper, hvor mengden data som analyserer trenger helt nye teknikker og teknologi (Gray & Szalay, 2007 ref. i Bell, Hey, & Szalay, 2009).

Ulike typer data er koblet til måten som informasjon er samlet inn på og behandlet. Data forekommer i mange ulike formater og filtyper, slik som tekstdokumenter, laboratorienotater,

feltnotater, kodebøker, lydfiler, bilder, video, gensekvenser, databaseinnhold, skript, algoritmer og operasjonsprosedyrer m.m.. Forskningsdata deles inn i ulike nivåer basert på hvor bearbeidet data er. Da omtales data for eksempel som rå (ubearbeidet), prosessert (tilrettelagt for analyser) eller publisert (Østbye et al., 2013).

Borgman (2015) problematiserer inndelingene. Hva som betraktes som rådata vil variere innen forskningsmiljøene. I praksis kan rådata være instrumentdata når instrumentet umiddelbart fanger opp et signal, eller når et bestemt fenomen forekommer under bestemte betingelser og registreres av instrumentet. I andre sammenhenger kan rådata være utfylte spørreskjemaer eller intervjuer fra forskningssubjektene.

Vaskede data legger grunnlaget for analyser, og det er fra disse konklusjoner trekkes. Avgjørelser om hvordan man skal behandle manglende svar «missing data», legge sammen variabler eller gjøre andre beslutninger for å tilrettelegge data kan være minimalt dokumentert. Dette kan ha en stor innvirkning på resultatene, fortolkningen, gjenbruk og etterprøvbarehet ((Blocker & Meng, 2011), ref. i Borgman (2015)).

Det er ikke bare datas behandling som er utfordrende for å klart definere typer data. Kombinasjonen av forskjellige datakilder kan føre til utfordringer for definisjoner og praksis. Kruse og Thestrup (2014) problematiserer skillet mellom primær forskningsdata, dvs. innsamlet av forskeren selv og sekundære forskningsdata, dvs. innsamlet av andre forskere eller offentlige data. Dette skillet blir utfordrende når data kombineres på nye måter eller fra flere kilder, slik at definisjonen kan ha betydning for eierskap og videre deling av data. Borgman (2015) fremhever at skillet mellom primær og sekundærkilder i humaniora er det nærmeste man finner rådata og vaskede data, slik som man gjør i realfagene og samfunnsvitenskapene. Hva som antas er primærdata er avhengig av konteksten, og ens utgangspunkt. Det som en person mener er sekundærdata kan være en annen forskers primærkilde.

Politiske retningslinjer og policydokumenter legger andre definisjoner av forskningsdata til grunn. OECD (2007) beskriver forskningsdata som:

“Research data are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings¹”. Definisjonen legger

¹ Det understrekes i rapporten at følgende elementer ikke inngår som forskningsdata: Laboratorierapporter, foreløpige analyser, utkast til vitenskapelige artikler, planer for fremtidig vitenskapelig forskning,

vekt på data fra primærkilder som anvendes i forbindelse med forskningsresultater, og valideringen av disse. Format og opprinnelse blir avgjørende for krav til arkivering av forskningsdata.

Norges forskningsråds retningslinjer legger vekt på skillet mellom ulike datatyper og formater ved krav om arkivering, slik at det omfatter digitale data som forskeren selv har produsert. Dette vises ved at forskningsdata defineres som «*registreringer/nedtegnelser/ rapporteringer i form av tall, tekster, bilder og lyder som genereres eller oppstår i forskningsprosjektet*»² (Norges forskningsråd, 2014b, s.10).

Arkiv- og biblioteksektoren har en ytterligere tilnæringsmåte til data. Borgman (2015) peker på at de mest konkrete definisjonene er knyttet til arkivenes praksisutøvelse. Likevel er det et paradoks fordi arkiver som lagrer og bevarer data ofte ikke er villig til å sette presise definisjoner, slik at arkivene kan tilpasse seg nye typer data etter hvert som de oppstår.

Arkivvirksomhet og bibliotekvitenskapen knytter definisjonen av forskningsdata til en prosessmodell. «Research data management (RDM) is about “the organisation of data, from its entry to the research cycle through the dissemination and archiving to valuable results” (Whyte & Tedds, 2011, s.1). Administrasjon av forskningsdata består av en rekke forskjellige aktiviteter og prosesser knyttet til datas livssyklus. Den består av design og produksjon av data, lagring, sikkerhet, bevaring, gjenfinning, deling og gjenbruk, hvor alle deler inkluderer tekniske muligheter, etiske betraktninger, juridiske aspekter og politiske retningslinjer.

De forskjellige definisjonene tyder på at databegrepet er ikke entydig. Denne studien fokuserer på forskernes praksis. I forskernes bevissthet vil inndeling av data etter metodisk fremgangsmåte og begrepsbruk stå i fokus. Samtidig som man har i bakhodet at forskningsfinansører og myndigheters krav til arkivering først og fremst gjelder digitale primærdata. I arbeidet med praktisk tilrettelegging av data for lagring og bevaring vil prosessmodellen for administrasjon av forskningsdata være sentral. Oppgaven vil ha behov for et helhetlig syn på forskningsdata, siden undersøkelsen berører kartlegging, synspunkter

fagfelle vurderinger eller personlige kommunikasjon med kollegaer eller fysiske gjenstander slik som laboratorieprøver, bakteriekulturer eller forsøksdyr.

² Kildedata, slik som kliniske data, værdedata eller tekstkorpus, faller utenfor definisjonen. Grunnen er at data er innsamlet eller bearbeidet av andre. Dermed utelukkes dokument/fysiske kilder og registerdata som omfattes av andre lovreguleringer.

og praksis hos forskere. I oppgaven vil det anvendes en definisjon av data som vektlegger bevisførsel for forskning. Dette er slik Borgman (2015) fremsetter data som “*representations of observations, objects or other entities used as evidence of phenomena for the purpose of research or scholarship*”.

Hun understreker at enheter blir data kun når noen bruker den som bevis for et fenomen, og en enhet kan være bevis for mange fenomen. På denne måten er ikke data rene eller naturlige objekter, men eksisterer kun i kontekst. Data vil ta deres betydning fra den spesifikke konteksten og fra personen som besitter data. Representasjonene varierer etter forsker, omstendighet og over tid. Dette utdyper hun ved å si at data er et middel for å komme frem til et vitenskapelig arbeid, vanligvis fremstilt som en artikkel, bok, konferansepresentasjon e.l. Hun mener at det å sette begrepet «data» på noe, er i seg selv en vitenskapelig handling. En del av det å bli forsker innebærer å kunne vurdere data. Dette betyr at forholdet mellom publikasjoner og data er sentralt, slik at data studeres innen rammeverket for vitenskapelig kommunikasjon.

Således, kan definisjonene på hva data er og hvordan data skal behandles oppsummeres med at det er avhengig av utgangspunktet og den rollen man har i arbeidet med forskningsdata. Dette kan tyde på mulige spenninger eller motsetningsfylte syn med tanke på aktivitetene i forskningssektoren. Derfor vil det videre belyses faktorer som kan virke inn på aktører i en forskningsinstitusjon, slik som forskningspolitiske føringer fra myndigheter og retningslinjer fra forskningsfinansiører. Andre påvirkningsfaktorer kan være vitenskapens verdier og normer på forskningsprosessens utførelse.

Forskningspolitiske føringer og retningslinjer for arbeid med forskningsdata

Myndighetene i Norge har valgt å følge opp retningslinjene fremstilt i «Principles and Guidelines for Access to Research Data from Public Funding» (Organisation for Economic Co-operation and Development, 2007). Retningslinjene omtaler at innovasjon innen forskning er nødvendig for å løse globale utfordringer som energi, helse, klima og naturressurser. Utveksling av ideer, kunnskap og data er grunnleggende for utvikling. Hurtig utvikling av datateknologi og internett fører til nye muligheter for anvendelser av forskningskildene, dvs. forskningsdata. Med dette utgangspunktet ønsker OECD at medlemslandene skal:

- Fremme en kultur for åpenhet, tilgang og deling av forskningsdata i forskningssamfunnet
- Øke oppmerksomheten omkring fordeler og ulemper ved tilgang eller deling av data, og eventuelle kostnader
- Vise behovet for regulering og etablering av praksis for datatilgang og deling i medlemslandenes forskningspolitikk
- Gi et felles rammeverk av forskningsdatatilgang (operasjonalisering)
- Gi anbefalinger om hvordan internasjonal forskningsdatadeling og formidling kan gjøres

Dette gir en idealisme om hvordan forskningen kan utvikle seg videre, og hvordan samfunnet vil nyte godt av forskningen. Forskningsdata som OECD henviser til, er utført av offentlige etater og/eller tildelt midler fra offentlige kilder. Å tilgjengeliggjøre faktainformasjon er en god forvaltning av offentlige investeringer, slik at tilgang til forskningsdata medfører en økt «return of investment» ved offentlige midler. Håpet er at det skal oppstå nye måter å koble tema for forskningen sammen, slik at nye områder belyses. Tanken er å si at «flere hoder tenker bedre enn ett». Ved tilgang på data, kan data brukes til å teste nye eller alternative hypoteser og analysemetoder. Dette vil være rasjonelt ved at data som er innsamlet kan resultere i flere analyser og meninger, og synliggjør nye mulige arbeidsområder for forskningen. På denne måten vil det kunne skapes sterke verdikjeder for innovasjon. Data kan kombineres fra flere kilder og gi ny kunnskap. Kunnskap som man ikke hadde muligheter til å få frem uten åpne data. Deling av data vil gi en større åpenhet omkring vitenskapelige undersøkelser, og forskningen styrkes gjennom muligheter for etterprøvbare. Publisering av data kan ha konsekvenser for den enkelte forsker. For forskeren vil deling av forskningsdata bidra til synliggjøring av forskeren. Dette mener OECD kan skape nye relasjoner og samarbeidsmuligheter.

Norske myndigheter har i etterkant av OECDs retningslinjer gjort en rekke grep for å synliggjøre forskningsdata som er viktig aspekt på den politiske agendaen. I forbindelse med innføringen av OECD retningslinjene i 2007, gjorde Norges forskningsråd en utredning til Kunnskapsdepartementet i 2008 om dagens praksis ved forskningsinstitusjonene (Norges forskningsråd, 2008). I Stortingsmelding nr. 30 (2008-2009) «Klima for forskning» fra Kunnskapsdepartementet (2009) vises det til at ”Åpen tilgang til forskningsdata og forskningsresultater er nødvendig for å øke tilgangen til eksisterende kunnskap og sikre samfunnet en uavhengig kunnskapsbase”. Det satses på en teknisk infrastruktur for

forskningsdata. Dette vises blant annet i «Veikart for norsk forskningsinfrastruktur», del I og del II (Norges forskningsråd, 2012a, 2012b). Budsjettinnspill viser hva som fagaktørene anser som viktig, blant annet vist i Norges forskningsråd og innspill til e-infrastruktur i «Store satsinger 2015» (Norges forskningsråd, 2013). Forskningsmeldingen «Lange linjer – kunnskap gir muligheter» (Kunnskapsdepartementet, 2013) legger også vekt på betydningen av infrastruktur, åpne tilganger og forskningsdata.

Myndighetene i Norge møter dermed de samme praktiske og teknologiske utfordringene i oppfølgingen av OECDs policy. Ekspertpaneler i EU har gitt anbefalinger om innføringen av en e-infrastruktur for forskning. Rapportene (Van der Graaf & Waaijers, 2011; Wood et al., 2010) peker på at tekniske løsninger for en e-infrastruktur er kun et element som det må tas stilling til. Andre spørsmål er:

- Utvikling av dataformater
- Adgangskontroll/endringsmuligheter av data
- Dokumentasjon av kontekst til data og kilden
- Hvem skal ta på seg kostnader?
- Hvilken intensiver skal anvendes (for at forskerne skal gjøre dette)?
- Hvor stor andel av data skal være gratis, og offentlig tilgjengelig?
- Hvordan skal koblinger mellom personopplysninger og data gjøres, og hvordan skal individets rettigheter ivaretas?

Før innføringen av en norsk policy for forskningsdata beskrev Thestrup et al. (2012) situasjonen i Norge som fraværende av en enhetlig nasjonal politikk, men dagens praksis reguleres fra ulike hold. En del av dagens praksis for forskning reguleres gjennom lovverket, slik som Personopplysningsloven (2000), Helseforskningsloven (2008) og Helseregisterloven (2001) og lover angående registerdata som Helseregisterloven (2001) eller Statistikkloven (1989).

I september 2014 lanserte Forskningsrådet sin policy for «Tilgjengeliggjøring av forskningsdata». Forskningsrådet etablerte sin policy i tråd med retningslinjene fra OECD og EUs policy³ (European Commission, 2013). Den norske policyen er at forskningsdata skal

³ EUs retningslinjer: "Regarding the digital research data generated in the action ('data'), the beneficiaries must:
(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:
(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
(ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan"

være åpent tilgjengelig, når data er helt eller delvis finansiert med offentlige midler. Data kan unndras tilgjengeliggjøring på grunnlag av sikkerhet, personvern, juridiske eller kommersielle forhold. Andre forhold kan komme i betraktning, men da må det ha store konsekvenser av økonomisk eller praktisk art. Dette betyr at forskerne er ikke forpliktet til å gjøre data åpent tilgjengelig. Dersom man ikke er villig til å gjøre data åpent tilgjengelig, så må det begrunnes.

Forskningsrådets policy blir supplert av retningslinjer (Norges forskningsråd, 2014b).

Retningslinjene gir flere detaljer omkring hvordan data bør behandles. I likhet med EUs retningslinjer ønsker Forskningsrådet at ved søknad om forskningsmidler skal det avgis en plan for databehandling, Data Management Plans [DMP]. Dette innebærer at forskeren må ta stilling til om deres data vurderes å ha verdi på lang sikt. Videre bør data lagres i sikre arkiver, med standardiserte metadata og ha internasjonale lisenser for bruk knyttet til seg. Når forskeren publiserer en vitenskapelig artikkel, så bør tilhørende data være publisert. Andre data skal gjøres tilgjengelig senest 3 år etter endt prosjekt. Dersom data ikke har langvarig verdi, så må tidspunkt for destruering etter en viss tid vurderes.

Forskningsrådet vil opprettholde krav i deres kontrakter om minimum 10 års arkivering av data. Dette er i tråd med tidligere etablert praksis på feltet hvor Forskningsrådets kontrakter innenfor samfunnsvitenskap, humaniora, helse, biofag, miljø og utvikling krever at forskningsdata overføres til arkivet ved Norsk Samfunnsvitenskapelig Datatjeneste [NSD]. I tillegg vil Forskningsrådet lage en oversikt over andre godkjente infrastrukturer for lagring, arkivering og tilgjengeliggjøring av data.

Persondata behandles ut fra krav satt av Datatilsynet. Personvernombud for forskning gir anbefalinger til hvordan man behandler, oppbevarer og lagrer data for å ivareta forskningsetiske normer og regler som de fleste forskningsinstitutter følger. Ved forskningsinstitusjonene er det varierende praksis for databehandling. Noen institusjoner har prosjekthåndbøker som beskriver retningslinjer og rutiner for oppbevaring og lagring av data, men de fleste har få eller ingen rutiner for lagring og tilgjengeliggjøring av data. Dette betyr at lagring av data reguleres av den enkelte forsker eller forskningsleder, dersom ikke Personopplysningsloven og Datatilsynet setter særlige krav. I de fleste tilfeller innebærer dette at data lagres slik at kun forskeren selv har tilgang til dem.

Oppsummert kan man si at forskningsdata reguleres gjennom forskningsfinansiering, retningslinjer hos de enkelte forskningsinstitusjonene, personvernombudet for forskning og forskjellige lover som regulerer forskning.

Reguleringer av forskningsdata i Norge er knyttet til internasjonale føringer, lovreguleringer og etiske normer og regler. Det antas at dersom forskning skal løse store samfunnsmessige og globale utfordringer forutsetter dette koblinger mellom forskning på nye måter. Utviklingen er nært knyttet til en teknologisk infrastruktur, hvor blant annet tilgangskontroll, lisensiering, åpen tilgang til data i arkivløsninger er sentralt.

De forskningspolitiske retningslinjene viser at kontekst, nytte og verdier gjenspeiles i omtale av forskningsdata. Forskningsrådet fokuserer på at det er digitale data som skal bevares.

Borgman (2015) fremhever at digitale data kan ikke fortolkes uten det tekniske apparatet som har vært med på å skape dem. Digitale data krever dokumentasjon, ikke bare for radene og kolonnene i regnearket, men også for prosedyrene som har vært brukt for å samle dem inn. Data har på denne måten ingen betydning, dersom man ser dem isolert. Data er nært knyttet til andre begreper som står sentralt ved formidling av vitenskap. Det er en glidende overgang til informasjonsbegrepet og kunnskapsbegrepet. I det videre vil derfor sammenhengen mellom begrepene data, informasjon og kunnskap belyses.

Data, informasjon og kunnskap

Informasjonsbegrepet er i likhet med databegrepet ikke entydig. Først vil det gis en presentasjon av informasjonsdefinisjoner, før informasjon vil eksemplifiseres inn på temaområdet forskningsdata gjennom OECD-policyen. Webster (2006) knytter informasjonsbegrepet til 6 faktorer som skaper en kvalitativ endring i informasjon som utgjør et kvalitativt sosialt system, informasjonssamfunnet. De 6 faktorene for informasjon kan defineres i lys av teknologi, økonomi, yrke, geografi, kultur eller informasjon som endringsagent hvor måten vi lever på og vårt handlingsmønster endres.

Teknologisk utvikling er et synlig tegn på endring som definerer en periode i tiden, og er dermed tegn på et informasjonssamfunn. Et økonomisk syn på informasjon innebærer at veksten i bruttonasjonalprodukt kobles til informasjonsaktiviteter slik at de kan sies å være en del av den økonomiske aktivitet. Porat (1977) skiller mellom primær og sekundær informasjonssektorer innen økonomien. Den primære er knyttet til en markedspris, mens den sekundære er mer indirekte, for eksempel informasjonsaktiviteter innen en organisasjon eller

stat, men som er viktig for deres organisering. Han beskriver en kopling mellom data og informasjon som en informasjonsaktivitet som kan knyttes til økonomiske mål og tallfestes: *“Information is data that have been organized and communicated. The information activity include all the resources consumed in producing, processing and distributing information goods and services”* (Porat, 1977, s.9).

En overvekt av stillingskategorier knyttet til informasjonsarbeid skapt gjennom yrker, utdanningsforløp og erfaring er ytterligere et karakteristika på informasjonssamfunn, i følge Webster (2006). Videre peker han på at informasjonsnettverk kan kobles til geografiske steder og konsekvensen det kan ha på organisering av tid og rom. I et “nettverkssamfunn” har man endret begreper omkring tid og rom, fordi klokken og avstand ikke lenger setter begrensninger slik at man kan ha relasjoner i nåtid uten at man fysisk møtes.

Den kulturelle forståelsen av informasjon viser til at samtidskulturen er informasjonsladet. Kulturen uttrykkes blant annet gjennom symboler. Symbolene kan formidles og mottas, gjennom media, personlig bekledning, sminke og kroppslig fremtoning. Den økte mengden beskjeder som sendes er et uttrykk for et informasjonssamfunn.

Det siste synet på informasjon fremhever at man kan ha mye informasjon, men dette utgjør ikke et informasjonssamfunn med mindre det endrer måtene som vi handler på. Det uttrykkes ofte som skillet mellom “having information” og “being informed”. Et informasjonssamfunn hvor teoretisk kunnskap skaper en kvalitativ endring som det tidligere har manglet omtales gjerne som kunnskapssamfunnet. Dette bygger på tanken omkring det “rasjonelle” mennesket, hvor man reflekterer om det sosiale og utøver selvrefleksjon, tar avgjørelser og planlegger ut fra den teoretiske kunnskapen som vi har om verden og som er formidlet til oss gjennom ulike informasjonskanaler.

Webster konkluderer med at informasjonssamfunnet er et vanskelig begrep fordi det er uklart hva som det består av og hvordan skille et informasjonssamfunn fra andre definisjoner av samfunn. Han fremhever at en forklaringsmodell ikke kan legges til grunn fordi endring skapes et av samvirke av sosiale, økonomiske, politiske og teknologiske faktorer.

Hvordan informasjon påvirker samfunnet beskrives av Braman (2006). Kategoriene er til dels overlappende med inndelingene til Webster (2006). Braman vektlegger i større grad

relasjonen mellom informasjon og kunnskap, og hvordan den får konsekvenser og påvirker samfunnet.

Hun forklarer at informasjon som ressurs forekommer kun når informasjon er en forutsetning for å kunne fungere, og når det er et innspill til en beslutningsprosess, produksjon eller byråkratisk prosess. Dette er ofte kvantitativ informasjon som antall bøker, e-poster osv. og sier ingenting om innhold, effekter, atferd eller nivå.

Informasjon som en vare, gjerne en informasjonstjeneste som kan kjøpes og selges. Dette er koblet til synet på informasjon som økonomi, som også forklares av Webster (2006).

Informasjon som «perception of pattern» innebærer at informasjon kan fortolkes i tid, hvor informasjonen har en fortid og en fremtid. Fortolkningen kan påvirkes av motiv, miljø- og årsaksfaktorer. Braman (2006) viser til et eksempel fra vitenskapen, hvor man skiller mellom data og informasjon, informasjon og kunnskap, kunnskap og klokskap. På denne måten får informasjon en inndeling i nivåer. Denne typen informasjon kan sies å ha en verdi knyttet til reduksjon av usikkerhet, og som kan forekomme ved bruk av teknologisk innovasjon, overgang til en annen sosial struktur eller ved introduksjon av en ny teori. Braman (2006) hevder at denne typen informasjon vises klart innenfor policyer ved støtte til grunnforskning og anvendt forskning og utdanning. Denne definisjonen er grunnleggende i en rekke policy prosesser, men kan være vanskelig å operasjonalisere for integrasjon i beslutningsprosessen.

Informasjon i seg selv kan få ting til å skje, slik at den er sin egen aktør, for eksempel gjennom automatisering av prosesser. Denne typen «Information as agent» er økende fordi programvare blir mer og mer komplisert, og sensorer inkluderes i flere ting.

«Information as a basin of possibility» knytter informasjon til sannsynlighetsteori.

Kvantitative data behandles ofte som fakta, men dette refererer egentlig kun til sannsynligheter. Kunnskap kan sees på som en samling med variabler, hvor sannsynligheten for at hendelser faktisk forekommer beregnes. Det bør i analyser tas høyde for at forskjellige sannsynlighetsutregninger kan forekomme ved forskjellige betingelser.

Den siste informasjonstypen er “Information as a constitutive force in society”.

Konstitusjonelle prinsipper beskriver et ideal, mens konstitusjonelle krefter kan sees empirisk ved at de kan føre samfunnet nærmere eller øke avstanden til det konstitusjonelle målet. Dette innebærer at informasjon er en mulighet for aktivt å forme konteksten (samfunnet).

Braman bemerker at det finnes ingen enkel definisjon på informasjon når den anvendes i policyutforming da informasjonsbegrepet vil inkludere verdier. De fleste europeiske myndigheter, inkludert Norge, legger OECDs policy angående forskningsdata til grunn i sitt eget arbeid, og viser til hvordan retningslinjene antas å ha innvirkning på forskningsaktivitet. I det videre vil informasjon og databegrepet eksemplifiseres gjennom en fremstilling av sentrale punkter i policyen.

I sin ytterste konsekvens kan det diskuteres om forskningsdata er informasjon, dersom man ser på informasjon som en kvalitativ endring som fører til et endret handlingsmønster, jf. Webster (2006). Dette kan man si fordi lagring og tilgjengeliggjøring av forskningsdata er i seg selv ikke nok for å kunne utnytte potensialet i data. Dokumentasjon er en forutsetning for at andre enn forskningsgruppen skal kunne tolke og anvende data, slik at en kvalitativ endring skal inntreffe. Imidlertid, så ser man at flere andre definisjoner av informasjon anvendes. Informasjonen betraktes som en ressurs ved at den er en grunnleggende forutsetning for at det skal forekomme endring i kunnskap (Braman, 2006). Informasjon blir i OECD-dokumentet behandlet som «information as perception of pattern» siden det forventes en nivåmessig endring over tid, for eksempel ved at man beveger seg fra data til kunnskap og innovasjon.

Informasjonens løsrivelse fra tid og rom, uavhengig av geografiske begrensninger, er en forutsetning (Webster, 2006). OECD-policyen fremhever internett som en ønskelig kommunikasjonskanal, da denne er globalt tilgjengelig uavhengig av tid og sted. Det forventes at lagring og deling av forskningsdata skal være med på å skape, og eventuelt løse, globale utfordringer som således kan ha innvirkning på sosiale, politiske, teknologiske og økonomiske faktorer.

Det økonomiske aspektet av informasjon (Braman, 2006; Webster, 2006) står sterkt i policyen siden det handler om å få størst mulig utbytte av offentlig finansiert forskning. Den økonomiske verdien av forskningsdata fremheves på forskjellige måter. Det snakkes om bedre «return of investment» for offentlige instanser som finansierer forskning ved at flere personer får tilgang til innsamlede data og kan benytte seg av disse. Likeledes snakkes det om

innovasjon, som kan legge grunnlaget for å skape nye verdier. Behovet for kostnadsanalyser for å vise verdien av forskningsdata, og deling av data som et grunnlag for å argumentere for midler til finansiering av datalagring diskuteres. Kostnader for å etablere en infrastruktur og lagre data vektlegges. Intensivsystemer kan være knyttet til økonomiske midler, men det kan også være andre faktorer slik som økt anerkjennelse.

Informasjonsbegrepet kan knyttes til yrkeskategorier som arbeider med informasjon (Porat, 1977). I OECD-dokumentet uttrykkes behov for yrkesmessige endringer eller satsinger innen sektoren. Det blir nevnt spesialiseringer innen områdene som tar for seg: Hvordan data kan tilrettelegges og brukes med vekt på tekniske aspekter, assistanse i forhold til de lovmessige reguleringer som kan oppstå, og støttespillere i forbindelse med å formidle informasjon omkring hva som er informasjonspolitikken. Behovet for å formalisere roller er påpekt, og organisasjonens behov for å tenke langsiktig omkring arbeidet med forskningsdata. Forskernes arbeidsoppgaver kan endre seg ved at krav omkring dokumentasjon av forskningsdata endrer seg, eller bruk av andre analysemetoder eller teoriområder. En konsekvens kan være endringer av arbeidsoppgaver med opprettelse av nye roller og stillingskategorier.

Et annet aspekt som forskningsdata forventes å påvirke er nettverk (Webster, 2006). Da menes nettverk definert som personlige relasjoner mellom forskere, hvor synliggjøring av tilgjengelige forskningsdata kan være et hjelpemiddel til å få nye kontakter og nettverk innenfor eget eller andre fagfelt. Samtidig omtales behovet for at teknologiske nettverk må interagere for at bruk av forskningsdata skal ivaretas. Det tenkes på muligheter for tilgang til lisensiert materiale, lage koblinger mellom ulike databaser gjennom felles standarder og overføringsprotokoller, publikasjoner og datasett for å skape forståelse for data og riktig anvendelse. Tilgjengeliggjøring av data på nett vil innebære en mulighet for synliggjøring av den enkelte forsker eller forskergruppe og forskningsinstitusjonen. Tilgjengeliggjøring av data kan brukes som et element i profilering av institusjonen gjennom benchmarking, eller ved tenkte målinger av gjennomslagskraft for datasett og kvalitet gjennom utvikling av «data set impact». Videre vil reguleringer av forskningsdata ha innvirkning på enkeltpersoners rettigheter med hensyn til opphavsrett, lisensiering og personvern, samt muligheter til å få tilgang til data.

Informasjonsbegrepet slik som det anvendes i “OECD Principles and Guidelines for Access to Research Data from Public Funding” legger ikke en bestemt oppfatning omkring informasjon

til grunn for forskningsdata. Policydokumentet tar for seg reguleringer knyttet til stadiene som informasjonen gjennomgår - fra opprettelsen, via innsamling, organisering, formidling og tilintetgjørelse. Den overordnede visjonen er å kunne løse samfunnsmessige, globale utfordringer gjennom bruk av teknologi og tilgjengeliggjøring av informasjon. Teknologi er et avgjørende hjelpemiddel, men samtidig et felt under utvikling. Forskningsdata som informasjon som knyttet til tid, kontekst og nivåer går igjen i flere ledd. Det økonomiske aspektet av informasjon er spesielt fremtredende i policyen, sammen med tema knyttet til juridiske aspekter ved regulering av forskningsdata. Policyen legger vekt på prinsipper omkring medbestemmelse fra mange parter og åpenhet i beslutningsprosesser. Retningslinjene berører både enkeltpersoner i samfunnet gjennom åpen tilgang, forskere som arbeidstakere og profesjonelle utøvere, organisasjoner som universiteter, datalagringsinstanser og organer som gir forskningsfinansiering og ivaretar personvern, samt myndighetene på nasjonalt nivå.

Forskningsdata har nære bånd til informasjon og kunnskap, men fortolkningen vil være avhengig av konteksten. Politiske retningslinjer har mål om å vise hvordan investeringer i forskning anvendes, og hvordan dette bidrar positivt til å løse store felles utfordringer i samfunnet. Retningslinjene vil ha konsekvenser for praksis ved forskningsinstitusjonene og for den enkelte forsker. Hvem skal håndheve reguleringer ovenfor forskningsmiljøene? Vil flere forskningsfinansiører eller forskningsinstitusjoner utvikle egne retningslinjer som følger de internasjonale føringene eller vil de velge andre løsninger? Vil myndigheter eller forskningsinstitusjoner innføre intensiver for data, og hvilke intensiver vil gis?

Retningslinjene i policyer henviser til myndigheters offentlige mål, men er likevel preget av normer og verdier. På et institusjonelt nivå i vitenskapen finnes egne normer og verdier som er med på å styre forskningsaktivitet uttrykt gjennom forskningens prinsipper. Dette legger grunnlag for å se nærmere på vitenskapelig aktivitet og praksis med verdier og normer som den enkelte forsker berøres av.

Vitenskapens verdier

Vitenskapens normer og verdier dannes og reproduseres gjennom utdanningsforløp og sosial praksis. Dette kan illustreres ved å se nærmere på vitenskapelig aktivitet, for tilslutt å se institusjonelle normer og verdier i tilknytning til OECDs retningslinjer for forskningsdata.

S. Kvale (1999) trekker frem to versjoner av det å utføre vitenskapelig aktivitet. Det første synet er knyttet til forskning som en prosess som følger formelle og eksplisitte standardiserte metoderegler. Dette er fremtredende i den eldre logisk positivistiske vitenskapsfilosofien. Det andre synet ser på vitenskap som en situasjonsbunden aktivitet, koblet til situert kunnskap og intuitiv ekspertise som ikke nødvendigvis følger analytiske og bevisste regler. Forskerens person, personlige kunnskap og troverdighet er sentrale faktorer ved forskningsprosessen.

I dag kombinerer utdanningssystemet to typer for læring, i følge Kvale. Dette er læring i forskerutdanningen som reflekterer formelle fremgangsmåter gjerne gjennom skriftlig vurdering. Den andre typen er læring som en sosial praksis som finner sted i relasjon med andre. Undervisning gjennom sosial praksis består ofte av overføring av ikke-verbal kommunikasjon og viderekomne forskere er rollemodeller. Faglig identitet skapes av ritualer under læretiden som viser fagets normer og verdier. Kunnskaper og selvevaluering foregår gjennom evalueringer av praksis. Evalueringer forsterkes gjennom erfarne veileders tilbakemeldinger om kunnskaper, men også om verdier, normer, selvbilde og forventninger. Den vitenskapelige aktiviteten kan knyttes til den formaliserte utdannelsen og praksisutøvelse for eksempel gjennom mesterlære, slik som Kvale peker på.

I sin fremstilling av praksisteori fremhever Cox (2012) at informasjonsvitenskapen trekker veksler på teoretikere som Lave og Wenger (1991) fra situert læringsteori /mesterlære eller Schatzki (2002) når praksisteori omtales. Han trekker frem praksisteorier som en teoretisk tilnærming som kan anvendes ved studier av forskningsmiljøer og informasjonsdeling. Pilerot (2013) fremhever at det ikke finnes en praksisteori, men flere måter å begrepsliggjøre og studere praksiser på. Praksis er knyttet til at aktører handler med forventning om en felles måte å gjøre ting på, slik at dette setter en praksis. Cox (2012) beskriver praksisteori som å sette fokus på stedet hvor aktiviteter utføres, for eksempel et forskningsmiljø. Videre er relasjonene mellom personene og ting som er involvert, deres handlinger og hvordan aktivitetene kommer til uttrykk sentralt. Aktørene befinner seg i et fellesskap, men kan deltakerne være ulike med hensyn til deres faglige bakgrunn, nasjonalitet, geografiske plassering, forskningsfokus, akademisk erfaring og meritter m.m. (Pilerot, 2013).

Informasjonsaktiviteter er integrert i de fleste praksiser, men det er sjeldent hovedfokus hos de sosiale aktørene (Cox, 2012). Ved praksisteori i informasjonsdeling peker Pilerot (2013) på at det er 3 dimensjoner som tas i betraktning:

1. Den kulturelle-diskursive dimensjon inkluderer normer og forventninger formet av sosiale og historiske forhold om hvordan man bør gå frem. Forskere forventes å anerkjenne deres bruk av andres ideer, slik som det gjøres gjennom den kulturelle handlingen sitering.
2. Den sosiale dimensjonen av praksis vektlegger hvordan personer interagerer, danner relasjoner, opplever identitet og tilhørighet.
3. Den materielle og økonomiske dimensjonen utgjør teknologi, fysisk rom med møblering og bygninger. Forskere kan forventes til å handle i tråd med økonomiske forventninger, eks. finansieringssystemer eller publiseringssystemet i akademien.

Cox peker på at individet er bærer av praksis, preget av en sosial kultur, men hun har en mulighet til å bryte med denne, på denne måten er det rom for endring og kontinuitet. Praksis kan bare forstås ut fra sin egen kontekst, slik at en praksis kan ha en annen betydning i et annet land, og kan bare forstås med en henvisning til de større sosiale strukturene, slik som økonomiske relasjonene eller kulturelle forhold. Tilnærmingen forutsetter å se på hvordan praksiser griper over i hverandre eller om det er konkurrerende praksiser. Dette kan skape utfordringer med hensyn til å vite hvor en praksis begynner og en annen stopper, fordi de fleste praksiser er integrert med hverandre.

S. Kvale (1999) fremholder at ved læring i praksisfelleskapet for forskere vil det være en overordnet interesse for å holde i gang en felles produksjon, samtidig som det vil internt være et dynamisk samspill og konkurranse med spenninger mellom forskjellige generasjoner, mellom tradisjon og fornyelse. Gjennom praksisfelleskapet læres det å sanse et fint differensiert forskningshierarki og forskeren må lære å håndtere de motstridende kravene mellom selvhevdelse og konkurranse og samarbeid.

Dette leder over til vitenskapens karakteristiske verdier og normer som reflekteres gjennom praksis. Dette beskrev Merton på 1940-tallet gjennom vitenskapens etos. Vitenskapens etos er et system av verdier og normer som oppfattes som bindende for forskere, og på denne måten styrer vitenskapen. Vitenskapens institusjonelle mål er å frembringe ny og veldokumentert empirisk kunnskap. Normene er bindende fordi forskerne oppfatter dem som effektive midler til målrealisering, dvs. oppnå ny vitenskap og de er moralsk forpliktende. Normene fungerer på institusjonsnivå, og overføres fra en generasjon til den neste via regler og forbilder og blir håndtert via sanksjoner (Gilje & Grimen, 1995). Mertons vitenskapssyn er positivistisk og forutsetter at vitenskapelige hypoteser skal være empirisk testbare. Ny kunnskap bygger

videre på gammel kunnskap slik at der er en kumulativ kunnskapsvekst. Vitenskapen utvikler seg via generaliseringer fra veldokumenterte fakta. Normene er nedfelt i vitenskapelig aktivitet og omtales som CUDOS-normene. Normene er kommunisme, (communism), universalisme (universalism), uegennytte (disinterestedness) og organisert skeptisisme (organized scepticism).

Kommunisme henviser til felleseie av vitenskapelige teorier og data. Å holde vitenskapelige oppdagelser for seg selv strider mot vitenskapens forpliktelse med å dele oppdagelsen med andre, slik at andre kan bygge videre på denne. Publikasjoner blir synonymt med å gjøre en vitenskapelig oppdagelse. Merton (1957) ser på offentliggjøring som en forutsetning for vitenskapelig fremskritt. Det finnes ikke en privat eiendomsrett til publiserte data. Forskerens eiendomsrett anerkjennes bare på den måten at en teori kan knyttes til hans navn. Normen impliserer at man ikke kan gjøre krav på data som er publisert, det er ikke mulig å ta patent på teorier eller kjøpe en vitenskapelig teori og publisere det under sitt eget navn. Merton (1957) viser til at anerkjennelse for originalitet er det sosialt validerte beviset på at man lever opp til de krav som ens rolle som vitenskapsmenn innebærer. Belønningssystemene i forskningen er i stor grad knyttet til anerkjennelse, slik som priser eller æresdoktorater.

Universalisme omhandler at teorier og data skal vurderes i lys av upersonlige kriterier. Personlige kriterier som for eksempel forskernes rase, kjønn, klasse, nasjonalitet o.l. er irrelevante faktorer når en teoris gyldighet vurderes. Dette er en viktig forutsetning for vitenskapelig utvikling og objektivitet.

Uegennyttighet er en norm knyttet til kontroll av tvilsomme personlige ytringer hos forskere for å skape sperrer mot tvilsomme teorier. Dette innebærer at særegne trekk som å kunne ta imot kritikk og muliggjøre repeterbare undersøkelser vektlegges i forskningsprosessen. Merton (1957) hevder personlig ærlighet er knyttet til den vitenskapelige etterprøvbareheten. Originalitet er en sentral verdi fordi den bidrar til å fremme kunnskap. Søken etter kunnskap er kulturelt definert som knyttet til søken etter sannhet. Dette kommer foran forskning som et levebrød. Merton peker på at de institusjonelle verdiene har svakheter ved at det legges for mye vekt på originalitet som et suksesskriterium. Å ikke lykkes med å skape originalitet kan oppleves som et pressmiddel for forskeren og bidra til å skape avvikende atferd som plagiering, falsifikasjon av data o.l.

Organisert skeptisisme omhandler muligheter til å reise kritiske spørsmål til etablerte rutiner og praksiser. Dette kan føre til konflikt med andre verdier og normer forskeren har, f. eks. politiske eller religiøse verdier.

Kritikken mot Merton er knyttet til at normene er allmenne og upresise. Normer og regler forteller ikke hvordan de skal brukes i konkrete situasjoner. Normene betyr ikke det samme for aktørene i ulike kontekster. Dette har resultert i en ambivalens i forskersamfunnet. Mitroff (1974) har fremsatt motnormer «antitester» som respons til Mertons normer for vitenskapens etos. Dette er en kritikk mot forskningens upersonlige og uavhengige karakter som vises gjennom universalisme. Motnormene viser til forskningens personlige karakter som kan føre til motstridende holdninger og atferd. Motnormene peker på at sosiale og psykologiske karakteristika ved forskeren er viktige faktorer for hvordan et arbeid vil vurderes, slik at hvem som legger frem en påstand har betydning i kraft av den personen som legger det frem. Hemmelighold er en nødvendig handling for å beskytte ens eget arbeid frem til publisering. Eget arbeid bør identifiseres ordentlig, slik at andre kan holdes ansvarlig for unøyaktigheter. Forskeren bør ha fullstendig tro på sitt eget arbeid, og tvile på andres.

Argumentasjonen fremsatt ved OECD-prinsippene henviser til en rekke av verdiene som fremsettes av Merton. Dette er blant annet offentliggjøring og publisering av data for at andre kan bygge videre på dette. Etterprøvbarhet av forskningen gjøres mulig for andre gjennom fremsetning av dokumentasjon, samt krav om anerkjennelse gjennom referering for å vise at gjenbruk av data bygger på tidligere forskning. Økonomisk vinning på data som er lagt ut, fremheves som en samfunnsmessig nytte. Uegennytte og anerkjennelse reflekterer Mertons normer, mens muligheter til fortjeneste for eksempel gjennom dagens teknologioverføringstjenester og patentering henviser i større grad til Mitroffs motnormer. For forskerne kan en konsekvens være at andre kan tjene penger på data som de har lagt ut, og dermed en fare for at andre får anerkjennelse for det som forskeren har gjort. Åpen tilgjengelighet kan føre til fare for mistolkning av data. I retningslinjene finnes en henvisning til at lagring, bevaring og deling av data krever dokumentasjon. Dette vil innebære kostnader både til bruk av tid til dokumentasjon, og oppbevaring og bevaring foretatt av forskerne og forskningsinstitusjon.

Databegrepet, slik som det forekommer i dag for den enkelte forsker/ forskningsgruppe, er spesifikt knyttet til forskningsaktiviteter gjerne satt inn en metodisk kontekst med vekt på

forskjellige metodiske innsamlingsmåter og analyse av data i form av rådata og ferdig analyserte, anonymiserte data. Krav fra forskningsfinansiører med hensyn til tilgjengeliggjøring av data gjennom arkivering krever dokumentasjon på fremgangsmåter for å muliggjøre gjenbruk av data. Gjennom tilretteleggingen av dokumentasjon og kobling med vitenskapelige presentasjoner vil dette være med på å omforme data til å være informasjon. Imidlertid, så pekes det på at dette er prosesser som krever mye. For forskeren vil dette være en endring av praksis fra tidligere som innebærer en annen måte å presentere faser i forskningsprosessen ved bruk av andres data og presentasjon av sine data. Plassering av forskning som knyttet til kontekst med tanke på tolkning, og bruk av verktøy for å utføre forskningen er sentralt. Innen vitenskapen pekes det på at resultater som fremkom på et punkt i tiden ikke nødvendigvis er gyldig til evig tid, eller at forskningen først blir verdifull etter som man får nye kunnskaper.

Praksisteori, vitenskapelige normer og motnormer viser til eksisterende spenninger innen vitenskapen som felt. Konteksten består av samarbeid og konkurranse. Verdier i forskning er todelt. Bidrag til fellesskapet gjennom å bringe ny kunnskap videre betraktes som en verdi i seg selv, i motsetning til å bringe forskning videre som en potensiell økonomiske faktor. Politisk argumentasjon for arkivering og deling av data reflekterer flere institusjonelle vitenskapelige normer. Nytteaspekter i form av å løse nye problemer og skape en økonomisk og samfunnsmessig utvikling står sterkt. Verdiene og normene reflekteres i undersøkelser omkring forskeres syn på lagring, bevaring og deling av forskningsdata.

Forskeres syn på lagring av forskningsdata

I løpet av de siste årene er flere studier på forskeres holdninger til lagring og deling av forskningsdata gjennomført. Undersøkelsene gjenspeiler kompleksiteten og de mange faktorene som virker inn ved arbeid med forskningsdata. Tekniske aspekter, sosiologiske og økonomiske faktorer trekkes frem som viktig i arbeid med forskningsdata. Studiene har ofte anbefalinger av praktiske tiltak (Andreoli-Versbach & Mueller-Langer, 2014; Norges forskningsråd, 2014a; Reilly et al., 2011; C. Tenopir et al., 2014; Carol Tenopir et al., 2011; Van den Eynden & Bishop, 2014).

De sosiologiske faktorene reflekterer elementer knyttet til arbeidsforhold og karakteristika ved den enkelte forsker. Dette kan være ansettelsesforhold, posisjon i forskningskollegiet

(status/rang), alder, samarbeid, opplevd ledelsesstøtte, holdninger til open science, frykt for misforståelser eller misbruk av forskning, sikkerhet for å ivareta personvern hensyn, tidsbruk ved tilretteleggelse av data og dokumentasjon og anerkjennelse gjennom sitering.

Dette vises blant annet av Carol Tenopir et al. (2011) som har sett på forskeres datadeling. Resultatene viste at viktige elementer for å dele data er skikkelige siteringsmuligheter, muligheter for samarbeid eller intensiver. Hindre for å gjøre data tilgjengelig er mangel på tid, finansiering, lagringsalternativer og manglende standarder. Frivillig deling av forskningsdata har en signifikant økning ut fra karakteristika ved den enkelte forskeren slik som stabile ansettelsesforhold, forskerens kvalitet, andelen publikasjoner i tidsskrifter med krav til datadeling og personlige holdninger til «open science» (Andreoli-Versbach & Mueller-Langer, 2014).

Resultatene av Norges forskningsråd (2014a) undersøkelse av 1 474 forskere viste at de fleste mener at det er viktig å dele forskningsdata, men i praksis er det ikke mange som gjør dette. De som har delt data er oftere mer etablerte forskere, slik at høyere alder gir en større sannsynlighet for at man har delt data. Deling av data gjøres oftest innen egen forskningsinstitusjon. Matematikk og naturvitenskap er fagfeltene som deler flest data. Undersøkelsen omfatter alle fagområder. Det er ikke grunnlag for å konkludere med at visse fagområder trenger flere tiltak enn andre. Skiller mellom fagområdene ble ikke funnet, men de peker på at ikke alle fagområder i undersøkelsen er representative eks. humaniora. Resultatene viser at viktige aspekter for forskerne er, blant annet: infrastruktur, sikker lagring, siteringer, retningslinjer/veiledninger, opplæring og ledelsestøtte/fokus.

Forskerne identifiserer en rekke barrierer for arkivering og deling av data. Tid for å tilrettelegge og dele data er den største barrieren, deretter følger manglende teknisk infrastruktur. Redsel for at deling av data kan redusere muligheten for fremtidige publikasjoner er forskerne i stor grad opptatt av. Frykt for at andre ikke forstår data, og tilslutt personvern/sikkerhet er også faktorer som virker inn. Å dele forskningsdata oppfattes av forskerne som et komplisert temaområde, hvor det mangler intensiver for deling av data (Norges forskningsråd, 2014a).

Videre peker de fleste forskere på at ledelsen ikke er spesielt opptatt av deling av data. Det pekes på at forskningsinstitusjonene i liten grad har veiledninger eller retningslinjer. Nesten

80 % mener de har liten eller ingen infrastruktur tilgjengelig. Likevel, så er de fleste forskerne tilfreds med dagens arkiveringsløsninger. Forskerne er spesielt opptatt av sikkerheten til data. De fleste forskere lagrer data hos seg selv eller på en institusjonell server.

De norske undersøkelsene bygger på spørreskjema fra PARSE insight (Kuipers, Hoeven, & Parse Insight, 2009). En tidligere undersøkelse av L. H. Kvale (2012) har sett på forskeres holdninger til deling av forskningsdata innen livsvitenskapene ved Universitetet for miljø og biovitenskap. Undersøkelsen viste at de fleste forskerne finner forskningsdata som blir gjenbrukt via sine kollegaer. Forskere som deler data mener at open science prinsipper er den viktigste årsaken til å gjøre dette. Enerett til første publisering og anerkjennelse ved gjenbruk er en forutsetning for deling av data. Holdningene til å gjøre data tilgjengelig er avhengig av hvor forskerne befinner seg i sitt karriereløp. Til tross for positive holdninger til deling av data finner hun at det er liten eller ingen forståelse blant forskerne av betydning av metadata. Ved sammenligning med Parse-studien finner hun at forskerne har høye forventninger omkring hvilken verdi deling av forskningsdata skal ha innen eget fagfelt enn andre forskere, samt at bruk av tidsskrifter og vitenskapelige forlags nettsider for å tilgjengeliggjøre forskningsdata er høyere enn på andre fagfelt.

Årsaken til at forskerne deler data er knyttet til mulighetene for ansettelse og opparbeidelse av anerkjennelse i fagmiljøet. Dette har Andreoli-Versbach og Mueller-Langer (2014) vist i et studium av praksis blant 488 forskere innen økonomi og business science. Undersøkelsen viser at 80 % ikke deler data, 16,8 % deler av og til, kun 2,46 % deler regelmessig. Deling av data kan, som nevnt, gjøres ved publisering. En studie av 141 anerkjente økonomiske tidsskrifter viste at de fleste tidsskrifter ikke har en policy for tilgjengeliggjøring av data, men et krav om arkivering av data (Vlaeminck, 2012). I forskning hvor data, kode og utregning er tilgjengelig er det varierende hvor tilrettelagt data er i praksis, slik at etterprøvbarehet kan gjennomføres. Policy med krav om kode og utregningsmetode er en «replication policy», men det er få tall av tidsskriftene som har en slik policy. Undersøkelsen viser at det er flere viktige intensiver for forskerne å dele data. Policy i tidsskrifter er en medvirkende årsak. Videre er forskernes mulighet for anerkjennelse når de deler data, og muligheter for å få støtte når man vil dele av data.

Undersøkelsen skiller mellom policy som inneholder et krav om å gjøre data tilgjengelig, og policy som inneholder informasjon om kode og utregningsmetode. Dersom ikke kode og utregningsmåte er med, så mener man at det er vanskelig å etterprøve resultatene.

Undersøkelsen fant at de fleste økonomiske tidsskriftene ikke har en policy for tilgjengeliggjøring av data. Blant studiene som har en policy, er det kun et fåtall av disse som har krav om en replication policy. Ved en nærmere undersøkelse av innholdet i policy, fant studien at de fleste tidsskriftene krever at forfatteren skal arkivere data. Imidlertid varierer det om man blir bedt om å gi mer detaljerte beskrivelser av data og kode.

Intensiver og investeringer er ofte fokus i de økonomiske faktorene. Kostnader ved lagring og tilrettelegging av data i arkiver trekkes frem. For å få finansiering er ofte en forutsetning at det lages en datahåndteringsplan eller data gjøres åpent tilgjengelig. Intensiver relateres til muligheter for å få økonomiske tilskudd på institusjonsnivå, dersom data publiseres av forskerene. Bruk av datapubliseringsordninger som en faktor ved ansettelse eller opprykksordninger er nevnt (Reilly et al., 2011; Van den Eynden & Bishop, 2014).

De tekniske aspektene gjenspeiler utfordringer ved arbeidsflyten mellom forskjellige tekniske systemer. Tilgjengeligheten av infrastrukturer som oppfyller bestemte tekniske krav og har tilstrekkelig IT-sikkerhet påpekes. Koblinger mellom ulike systemer som inneholder publikasjoner og data med reguleringer gjennom standarder som lenker og lisenser er omtalt.

Internasjonalt er det identifisert funksjonalitet og støttetjenester som kan benyttes i arkiver for forskningsdata (Dillo & De Leeuw, 2014; PARADE, 2009; Reilly et al., 2011). Samlet utgjør funksjonaliteten og støttetjenester kriterier på en kvalitetsmessig standard som dataarkiver vurderes etter før man sier at de tilbyr gode tjenester for lagring og bevaring av forskningsdata. De nevnte rapportene beskriver samlet en rekke funksjonalitet som kan anvendes i praksis på infrastruktur i arkiver.

PARADE White Paper (2009) fremhever den tekniske strukturen for avlevering, lagring og deling av forskningsdata som de mener vitenskapelig e-infrastruktur bør inkludere. Det nevnes blant annet: Åpen lagringsmulighet slik at brukeren kan lagre data lett. Permanente lenker som gir arkivene muligheter til å følge data tilbake til deres opprinnelse og beskriver karakteristika ved data. Metadataregistrering slik at data effektivt kan administreres, brukes og forstås. Interoperabilitet mellom systemer, både format og utveksling av informasjon i

registreringsfelt. Regulering av tilgang og lisenser for data. Lagringsmåter som sikrer at data kan anvendes over et bestemt tidsrom. Høy pålitelighet slik at forskeren kan stole på tilgjengeligheten av data.

I den overnevnte rapporten legges det stor vekt på den Flere nyere rapporter og policy initiativer legger vekt på muligheter for gjenbruk av data. Wood et al. (2010) tar for seg forventede fordeler med lagring og deling av data. Denne rapporten følges opp av “ Van der Graaf og Waaijers (2011) som ser på praktiske aspekter for å kunne gjennomføre deling av forskningsdata. En nyere rapport tar i større grad for seg koblinger mellom data og publikasjoner Reilly et al. (2011). Her understrekes betydningen av referanser for datasett med permanente lenker. Å bli kreditert for sine data er viktig for forskeren.

Informasjonselementer som inngår i referanser basert på standarder for siteringspraksis for datasett er det viktig for arkivene å inkludere.

Både norske og internasjonale undersøkelser peker på de samme faktorene som virker inn på deling av data. Personlige karakteristika ved forskeren, samt forhold i arbeidssituasjonen som bruk av tid og anerkjennelse for forskningsarbeidet og muligheter til første publisering.

Videre er det en rekke hindre som er koblet til arbeidssituasjonen som føringer, ledelsesstøtte og praksis i forskningsmiljøet, men andre strukturelle forhold som infrastruktur, veiledninger og retningslinjer, krav og intensiver virker inn.

De fleste undersøkelsene peker på tiltak for å øke deling av forskning. Hva skal til for at forskere skal være motivert for å dele forskningsdata?

Anbefalingene for tiltak slik at forskere vil arkivere og dele forskningsdata har en stor bredde. Tiltakene reflekterer den anvendte tredelingen i sosiologiske, tekniske og økonomiske faktorer. For å bøte på mangler ved sosiologiske forhold fremheves etablering av best practice diskusjoner innen et fagmiljø, opplæring og støttetjenester. Mer formelle reguleringer som innføring av institusjonelle retningslinjer med krav til publisering av data og bruk av datapublisering er foreslått. Ivaretagelse av anerkjennelse for forskeren ved at det finnes siteringsstandarder for data på lik linje som ved publikasjoner (Norges forskningsråd, 2014a; Reilly et al., 2011; Carol Tenopir et al., 2011; Van den Eynden & Bishop, 2014).

De økonomiske faktorene er forslag til etablering av infrastrukturer, og intensiver ovenfor forskningsinstitusjonene.

De tekniske faktorene kan avhjelpest med en rekke mulige tiltak, slik som støttetjenester, innføring av tekniske standarder, registrering og søkbarhet i datakataloger.

Tiltakene som Norges forskningsråd (2014a) nevner er blant annet informasjon om fordeler og ulemper ved å dele forskningsdata, opplæring i administrasjon av forskningsdata og metadata. Det bør skaffes bedre informasjon om hva som trengs for at infrastrukturen blir bedre. Analyser med henblikk for å fastsette behov er viktig, da investeringer i infrastruktur er kostbart. Retningslinjer og krav ovenfor forskerne har blitt gjort gjennom opprettelsen av en policy for tilgjengeliggjøring av forskningsdata for prosjekter som Forskningsrådet selv finansierer. Dette er kun en del av forskningen i Norge, slik at de anbefaler at institusjonene blant annet oppretter egne datahåndteringsplaner (Norges forskningsråd, 2014b). En oppsummerende figur for tiltak fra Norges forskningsråd (2014a) kan finnes i appendiks A.

Et av insentivene for å tilgjengeliggjøre forskningsdata er knyttet til at forskeren får anerkjennelse for det arbeidet som er gjort. En av måtene dette kan gjøres på er gjennom å lage koblinger mellom publikasjoner og data. “Integration of Data and Publication: Opportunities for data exchange” (Reilly et al., 2011) viser til fem krav fra forskerne for å lage koblinger. Dette er behov for å plassere data i et arkiv og gjøre data sikker for gjenbruk, behov for å gi tilgangsreguleringer for bruk av data, muligheten til å integrere data og publikasjoner på en bra måte, få uttelling for data som et førsteklasses forskningsobjekt og få dekket kostnader ved datalagring og gjenbruk av data.

Van den Eynden og Bishop (2014) har gjennom en oppsummering av flere forskningsprosjekter belyst mulige tiltak for å øke åpen tilgjengelighet av forskningsdata:

- Undervisning i forskningsmetode bør inkludere opplæring i hvordan forskningsdata kan deles. Dette er et viktig tiltak for at deling av data skal bli en del av forskningspraksisen.
- Forskningsmiljøene må selv etablere en forventning i miljøet om åpenhet, og hva som de anser som «best practice» på eget fagfelt.
- Forskningsinstitusjoner og finansører kan kreve en plan for hvordan data fra prosjekter skal behandles, datahåndteringsplaner.
- Det bør gis finansiering av kostnadene ved tilrettelegging og publisering av data.
- Forskningsdata kan bli en del av vurderingskriteriene ved ansettelse eller forfremmelser.
- Det må være tilgjengelige rådgivningstjenester for hvordan tilrettelegging for lagring og deling av data kan gjøres.

- Det bør være gode, stabile koblinger mellom publikasjoner og tilhørende data.
- Arkiver for forskningsdata bør finnes i registre for forskningsdata, og det bør være mulig å søke opp datasett integrert i litteraturdatabaser/bibliotekataloger.

Oppsummert vil dette si at det ikke bare er et enkelt felt som medvirker til at forskerne vil lagre, bevare og/eller dele data. Økt deling av data betraktes som en felles innsats blant mange aktører innen forskningen, slik som forskningsfinansiører, forskergrupper, forskningsinstitusjoner, forlag og dataarkiver. Bedre infrastruktur henviser til tekniske aspekter, sikkerhet og ivaretagelse av personvern henviser til juridiske aspekter, og sosiologiske faktorer knyttes til tidsbruk til klargjøring av data, sjansen for at andre kan misforstå innholdet i data, manglende anerkjennelse for arbeidet som er gjort med forberedelse og innsamling av data.

Forskningsdata i praksisfeltet

De mange definisjonene og faktorer som påvirker forskningsdata gjør det i mange tilfeller lettere å forstå forskningsdata gjennom arbeidsoppgaver. Prosessen for «Research Data Management»[RDM] er illustrert i beskrivelsen av 5 faser som forskningsdata gjennomgår fra begynnelsen til slutt i forskningsprosessen (Jones et al., 2013):

1. Ved oppstarten av et forskningsprosjekt vil forskeren opprette og skrive en datahåndteringsplan «Data Management Plan» [DMP]. Dette er et krav fra utdanningsinstitusjonen eller forskningsfinansiører, og gjøres ofte ved søknad om forskningsmidler.
2. Under selve forskningsprosessen berøres forskerens håndtering av data omtalt som «Managing Active Data».
3. Rett etter forskningsprosjektets avslutning blir det gjort en utvelgelse av data som skal tas vare på videre, «Data Selection and Handover».
4. Lagring av data i arkivløsninger, «Data Repositories», gjøres etter forskningsprosjektet. Data tilrettelegges for langtidslagring.
5. Deling av de data som er tatt vare på bør formidles. Dette kan gjøres gjennom datakataloger «Data Catalogues». Dette er en hjelp til at andre forskere kan finne tilbake til data, og data kan gjenbrukes.

I hele prosessen vil flere aktører være involvert, og hver av de 5 fasene kan bestå av mange aktiviteter. Det pekes på at universitetene og høyskolene utgjør en viktig kontekst og støttefunksjon for forskeren.

Storbritannia, USA, Australia og Nederland regnes for å være langt fremme i deres arbeid med forskningsdata (Larsen et al., 2014). En tidligere studie av status for forskningsdata i Danmark har vist at utviklingen av infrastruktur og tjenester for RDM befant seg i en tidlig fase (Van der Graaf & Waaijers, 2011). Status for forskningsdata i Norge ikke kommet så langt (Thestrup et al., 2012). Per dags dato er det ingen høyere utdanningsinstitusjoner i Norge som tilbyr tjenester innen alle de 5 fasene av RDM som Jones et al. (2013) beskriver. Krav om utforming av datahåndteringsplaner er ikke en praksis som er etablert. Høyere utdanningsinstitusjoner i Norge tilbyr noen tjenester innen forskningsdata, og flere tjenester er under planlegging.

I det videre vil det gi en kortfattet beskrivelse av prosessen som involverer forskningsdata med utgangspunkt i forskerens standpunkt satt inn i en norsk kontekst, illustrert gjennom mine figurer⁴ i appendiks E.

Ved prosjektetablering arbeides det med å planlegge forskningsprosjektet og finne samarbeidspartnere. Forskeren/forskergruppen må ta stilling til om de selv skal samle inn data, bruke andres data eller registerdata. I mange tilfeller søkes det etter forskningsmidler, og etableres avtaler som regulerer rettigheter og samarbeid. Datahåndteringsplaner er en del av søknaden som utarbeides. I dag berøres forskerne mest av retningslinjene til forskningsfinansisørene, Norges forskningsråd og EU. Arbeid med policy for RDM ved universiteter og høyskoler er startet, men mange institusjoner har ikke klare retningslinjer ut over det som er regulert gjennom lovverket og forskningsetiske normer. Videre er det aktuelt å skaffe seg nødvendige godkjenninger, hvor vurderinger av databehandling og datasikkerhet kan inngå. Alle norske forskningsinstitusjoner er forpliktet til å ha et personvernombud for forskning. Personvernombudet ivaretar personvern og etiske utfordringer i forbindelse med

⁴ Om modellen/figuren. Det er ikke tatt med alle detaljer for alle varianter av forskningsprosessen, siden dette kan variere noe med hensyn til bruk av forskningsmetode for å gjennomføre studier. Målet er at figuren skal gi et innblikk i noen av datarelaterte aktiviteter som forekommer i forskningsprosessen. Noen av prosessene foregår samtidig i tid, og prosessen kan derfor være ikke-linær. Modellen er inspirert av et «økosystem for forskning» fremsatt av CRISTin.

forskningsprosjekter, og kontrollerer lovpålagt beskyttelse av forskningssubjekter. De fleste høyere utdanningsinstitusjoner og forskningsinstitutter benytter Norsk Samfunnsvitenskapelig datatjeneste [NSD] som personvernombud. Internasjonalt reguleres forskningsetiske aspekter i stor grad gjennom forskningsfinansiører eller den enkelte utdanningsinstitusjon, slik som det for eksempel fremkommer i en datahåndteringsplan.

Ved prosjektgjennomføring vil forskeren velge publiseringskanaler, hvor opphavsrett til data eller publisering av data tilhørende en artikkel kan være et krav fra forlaget. Alternativt kan forskeren velge å publisere materiale i egne datatidsskrifter. Forskningsaktiviteter med henhold til analyse og bearbeiding av data kan inkludere utvikling av dokumentasjon som variabelbeskrivelser og kodebøker av studiet.

Prosjektet beveger seg inn en avslutningsfase. Etter at forskerne har gjort sin første publikasjon på data vurderes deres bevaringsverdi og eventuell lagring over kortere eller lengre tid foretas. Forskerne vil i denne fasen ta stilling til om data skal lagres i et dataarkiv, selv om dette kan gjøres ved oppstarten av prosjektet. Forskeren kan benytte flere ulike typer arkiver for lagring av forskningsdata. De kan forekomme i tilknytning til en institusjon, eller på nasjonalt eller internasjonalt nivå. Arkivene kan være knyttet til et eller flere faglige forskningsområder eller knyttet til bestemte forskningsmetoder. Langtidslagring innebærer en tilpasning av data til ny programvare, hvor eventuelt gjenbruk av data reguleres av lisensavtaler som forskeren selv har satt på data.

Fra et institusjonelt standpunkt vil det i større grad være fokus på ansvarsområder og rollefordeling mellom aktørene i RDM-prosessen. Dette kan illustreres gjennom UiOs forslag til institusjonell handlingsplan for forskningsdata⁵, som viser en inndeling i ansvarsområder og roller. I deres modell skilles det mellom 4 nivåer hvor interaksjonen mellom forsker og tekniske systemer beskrives, slik at ansvarsområder stadfestes. Tydelige likheter med de 5 fasene for forskningsprosessen (Jones et al., 2013), men uten fokus på DMP.

Nivå 1: Ved prosjektstart er den tekniske infrastruktur knyttet til det enkelte forskningsprosjekts behov, slik at ansvaret ligger til forskerne selv/forskningsgruppen. Nivå 2: Deretter skisserer rapporten et behov for et institusjonelt ansvar og løsning, som er knyttet til deling av data internt mellom deltakere i forskningsprosjekter, både ved egen institusjon og

⁵ Rapporten er utformet av en arbeidsgruppe ved Universitetet i Oslo (2015), med hensikt til å legge et grunnlag for et vedtak for videre arbeid med forskningsdata. Forslagene er ikke vedtatt per dags dato.

prosjektdeltakere utenfor institusjonen. På dette nivået vises det til et behov for midlertidig lagring av data som det er knyttet restriksjoner til, det vil si begrensning som følge av lovmessige reguleringer. Dette kan være før eller etter det er tatt stilling til om data skal lagres over lengre tid. Forslaget fremstiller dette som et institusjonelt ansvar, ev. samarbeid med andre aktører i universitets- og høyskolesektoren. Nivå 3: Lagring av endelige forskningsdata i nasjonale eller internasjonale dataarkiv, hvor det vises til at ansvaret og løsninger kan finnes nasjonalt, på programnivå eller institusjonelt. Nivå 4: Tilslutt pekes det på utnyttelse av andres forskningsdata gjennom bruk av datakataloger.

Utfordringer for forskningsinstitusjonene

Arbeidet med forskningsdata gjennom hele forskningsprosessen vil være en ny virksomhet for mange høyere utdanningsinstitusjoner. Utdanningsinstitusjonene må ta stilling til om hvilken politikk de ønsker å føre for datahåndtering, og hvilken praktisk infrastruktur som forskerne kan benytte seg av og hvilke oppgaver som institusjonen selv skal ivareta og eventuell rollefordeling innen egen institusjon. Forskningsinstitusjoner kan finne anbefalinger om aktiviteter som bør på plass, bl.a. i League of European Research Universities (2013) [LERU] Roadmap for Research Data.

Et første steg er å undersøke behovene ved egen institusjon for å identifisere hvor det er nyttig å sette inn arbeidet, og eventuelt lage en egen forretningsplan for arbeidet (Jones et al., 2013). Årsaken til at dette er viktig er at forskjellige fagområder kan ha ulike behov på grunn av metoder, standarder og dataformater som benyttes av forskerne (Thestrup et al., 2012).

En systematisk tilnærming til forskningsdata ved en forskningsinstitusjon anbefales å starte med en kartlegging av forskernes behov og status ved forskningsinstitusjonen (League of European Research Universities, 2013). Det vises blant annet til to verktøy anvendt blant annet i Storbritannia. Det foretatt en rekke casestudier ved universitetene basert på to ulike verktøy Data Asset Framework [DAF] (Data Asset Framework, 2009) og Collaborative Assessment of Research Data Infrastructures and Objectives [CARDIO] (Digital Curation Center). CARDIO er et verktøy strategisk policyutvikling av forskningsdata innen en forskningsinstitusjon eller forskningsgruppe. Dette verktøyet brukes i stor grad for å sammenligne eksisterende policy i ulike deler av en organisasjon og etablere felles

retningslinjer. Handelshøyskolen BI har ikke for øyeblikket retningslinjer for forskningsdata ved den enkelte instituttene eller en felles policy for forskningsdata, slik at et bruk av verktøyet vil være lite hensiktsmessig.

Data Asset Framework

Data Asset Framework⁶ ble opprinnelig startet som et prosjekt finansiert av JISC. Det var en del av Humanities Advanced Technology & Information Institute [HATII] og Digital Curation Center. Data Asset Framework har blitt anvendt av en rekke universiteter i Storbritannia: University of Bath, University of Edinburgh, University of Exeter, University of Northampton, University of Southampton, Imperial College og University of Oxford. Etter prosjektet har DAF blitt tilpasset av andre universitet, for eksempel ved Georgia Tech i USA.

Data Asset Framework kartlegger forskernes praksis for datalagring og holdninger. Resultatene brukes som et grunnlag for å forbedre eksisterende praksis. Undersøkelsene kan foretas på flere nivåer i organisasjonen, dvs. for en enkelt forskningsgruppe, et institutt eller for hele institusjonen. Målgruppen er avhengig av hensikten med undersøkelsen. Dette bestemmes ved at planleggingen må tilpasses den enkelte institusjon, hvor man spesifiserer detaljnivå på informasjonen man ønsker å finne i samarbeid med partnere i institusjonen. Målet med undersøkelsen er å kunne uttale seg om hvordan forskere ved en forskningsinstitusjon forholder seg til forskningsdata.

Kort beskrivelse av Handelshøyskolen BI

I lys av forslagene fra Forskningsrådet er det ønskelig med en større bredde på tilbakemeldingene fra forskerne. Det er hensiktsmessig å finne ut hvilket tilbud forskernes har behov for. Resultater kan være med på å legge grunnlaget for en eventuell politikk for forskningsdata. I denne undersøkelsen vil det tas et institusjonelt utgangspunkt, hvor forskernes behov er i fokus for undersøkelsen. Målgruppen for undersøkelsen er forskere ved den private stiftelsen Handelshøyskolen BI. Fagområdene ved Handelshøyskolen BI spenner vidt, men mye av forskningen er knyttet til fagområdene finans, økonomi og arbeids- og organisasjonsfag. Handelshøyskolen tilbyr utdanning på alle nivå innen høyere utdanning, fra bachelor til doktorgrad. Det er 4 studiesteder ved Handelshøyskolen BI hvor forskerne er fysisk lokalisert: Oslo, Stavanger, Bergen og Trondheim. Handelshøyskolen BI består av 8

⁶ Navnet var Data Audit Framework, men tilbakemeldingene i løpet av prosjektet innebar en misnøye med beskrivelsen som en revisjon ("audit") og dermed ble fokuset på verdsette ressurser ("asset").

institutter, hvor det er en rekke forskningssentre knyttet til instituttene. Organisatorisk sorterer alle institutter under «Forskning og fagressurser». Alle forskerne har tilhørighet til ett av de 8 instituttene ved BI.

Arbeid med infrastruktur for forskning har tidligere vært berørt i en rekke sammenhenger ved Handelshøyskolen BI. Tidligere arbeider som berører feltet forskningsdata har vært gjort, slik at noe intern dokumentasjon finnes allerede.

Dokumentasjon av infrastruktur for forskning er et internt dokument ved Handelshøyskolen BI. I løpet av august-oktober 2010 ble intervjuer med representanter fra instituttene ved Handelshøyskolen BI gjennomført. Undersøkelsen ble gjort i forbindelse med tildeling av økonomiske midler. Infrastruktur ble vidt definert. Fokus var på faktorer viktige for utføring av forskning på flere nivåer⁷. Det ble stilt åpne spørsmål om datainnhenting og databehandling hvor representanter fra instituttene kunne uttale seg blant annet om hvilken type primærdata benytter forskere og hvordan får forskerne tilgang til dataene. Øvrige spørsmål var knyttet til behov for litteraturressurser og administrativ støtte til forskning.

Forskere ved Handelshøyskolen BI henter i stor grad inn data selv, altså samler inn primærdata i forbindelse med at det utføres brukerundersøkelser og intervjuer. Registerdata fra for eksempel Statistisk sentralbyrå eller Skattedirektoratet. Noen av forskerne anvender informasjon fra databaser med selskapsinformasjon eller litteratur kilder som utgangspunkt for forskningen. Mesteparten av informasjonen analyseres i programvare som er vanlig ved forskningsinstitusjoner, slik som SPSS, STATA eller NVIVO. Til innsamlingen kan det benyttes nettbaserte spørreskjemaverktøy tilgjengelig ved Handelshøyskolen BI. En del av forskningsmiljøene lager egne databaser med informasjonen som de har samlet inn gjennom flere år. Men de fleste forskerne oppbevarer filer på sine egne datamaskiner. Det deles i relativt liten grad, ofte er det masterstudenter som benyttet forskernes data.

Bakgrunns materialet antyder hvilke data forskerne benytter seg av, men en relativt liten gruppe personer er blitt rådspurt. Det ønskes et overblikk over forskernes praksis ved bruk av forskningsdata for hele organisasjonen. Ved en spørreundersøkelse kan man få svar fra flest mulig forskere, og se om det er noen datatyper eller formater som er mer brukt ved forskningsinstitusjonen. Arbeidsmiljøet på BI består av forskere med ulik nasjonal bakgrunn, slik at survey undersøkelsen vil holdes på engelsk for å få svar fra flest mulig forskere.

⁷ I denne oppsummeringen er det kun tatt ut momenter som berører forskningsdata som tema.

En konsekvens av samtalene som tidligere har ført til et påbegynt arbeid med å få lage en arkivløsning for forskningsdata. Arkivet BIRD, BI Research Data, foreligger i en pilotløsning, hvor det kommet innspill om behov, slik som sikkerhet, brukervennlig grensesnitt m.m. Arkivet vil inneholde rådata. Målet er at arkivet skal følge internasjonale standarder.

Metode

I dette kapittelet beskrives bakgrunnen og valg av metode for undersøkelsen som er utført. Det blir gjort rede for kvaliteten i studiet og vurderinger av metodens egnethet og etiske betraktninger.

Valg av metode

Det samfunnsvitenskapelige perspektiv er utgangspunktet for undersøkelsen. Gjennom forskningen prøver man å dokumentere en virkelighet, helst så korrekt som mulig. De ulike elementene som antas å påvirke observasjoner og behandling av materiale ønsker man å ta hensyn til dette i utforming av spørreskjema og analyser.

«Spørreskjemaundersøkelser er en betegnelse for en gruppe metoder for strukturert datainnsamling ved hjelp av spørsmål og svar» (Østbye et al., 2013, s. 135). Målet er å skaffe data som kan brukes til å klassifisere og telle opp personer som har kombinasjoner av bestemte kjennetegn, slik at dette blir grunnlaget for en kvantitativ analyse. Undersøkelsen følger den empiriske forskningstradisjonen som anvender tallmateriale for å beskrive et fenomen, i dette tilfellet hvordan forskere forholder seg til forskningsdata. Hensikten med en spørreundersøkelse kan være å kartlegge adferd, måle kunnskaper og/ eller meninger. I undersøkelsen legges det vekt på kartlegge hvordan forskere anvender og tenker omkring forskningsdata, samt deres meninger omkring hvilke behov de selv har for videre kompetanseutvikling. Forskeren har mulighet til å gi innspill til hvordan de tenker arbeidsplassen deres bør utvikle seg videre for å kunne støtte dem i deres arbeid med forskningsdata.

Utvalget

Den totale populasjonen av forskere i Norge er knyttet til høyskoler og universiteter i Norge, forskningsinstitutter eller helseforetakene. Tilstandsrapporten for høyere utdanning 2014 (Kunnskapsdepartementet, 2014) viser at det ble utført 33 006 årsverk i universitets og høyskolesektoren i 2013. Forskere utgjør de faglige stillingene hvor dette utgjorde 59 % av årsverkene, dvs. 19474 stillinger. Stillingen er fordelt ut over universitets- og

høyskolesektoren. I alt finnes det 56 universiteter og høyskoler, hvor Handelshøyskolen BI er en av disse. Tall fra DBH viser at andelen faglige tilknyttede var 325 i 2014, fordelt på 239 vitenskapelige ansatte og 86 doktorgradsstudenter.

I denne undersøkelsen er det ikke foretatt en randomisering av populasjonen til å utgjøre et utvalg. Utvalget vil bestå av alle vitenskapelige ansatte og stipendiater på BI. Dette utvalget vil være et bekvemmelighetsutvalg, da man benytter de som ansatte som man har mulighet til å nå gjennom deres arbeidsplass. Utvalget har en viss skjevhet i forhold til den totale populasjonen av forskere i Norge, ved at det er overvekt av forskere med økonomisk-administrativ bakgrunn ansatt ved Handelshøyskolen BI.

Deltakelse i undersøkelsen er knyttet til respondentens egen villighet til å svare, slik at det gjøres en selvseleksjon. utfordringer med selvseleksjon er knyttet til responsrate og mulige skjevheter med hensyn til hvilke ansatte som svarer. Undersøkelsesdesignet, som benyttes, er et deskriptivt tverrsnittsstudie, dvs. en undersøkelse som beskriver et bestemt tidspunkt.

Det kvantitative undersøkelsesdesignet; spørreundersøkelsen

Målet med undersøkelsen var å belyse forskernes praksis med forskningsdata, få innspill til kompetansebehov og være med å legge et kunnskapsgrunnlag for videre utvikling av retningslinjer for forskningsdata på BI. Studien hadde som hensikt å gi:

- Bedre kunnskap om forskernes praksis i dag med hensyn til datalagring og deling.
- Muligheter til å se om enkelte institutter eller ansattegrupper har samme behov for infrastruktur til lagring og deling av data. Forskjellige fagområder kan ha ulike behov på grunn av metoder, standarder og dataformater som benyttes av forskerne.
- Innsikt i forskernes kompetansebehov/ønsker om opplæringstiltak.
- Bakgrunnsmateriale om forskningsdata som kan brukes inn mot BIs strategi og planer for arbeid med forskningsdata.

Kunnskap om forskernes praksis kan legge grunnlag for å si noe om hvilke eksterne nasjonale tjenester som forskerne kan benytte, og identifisere eventuelle tilbud som bør finnes lokalt. Kunnskapen kan legge grunnlag for å finne mulige samarbeidsrelasjoner blant andre norske aktører.

Undersøkellesobjektene er forskere ved Handelshøyskolen BI. Analysen benytter seg av primærdata, i og med at informasjonen samles inn særskilt for denne undersøkelsen.

Beskrivelse av spørreundersøkelsen

Spørreskjemaundersøkelsen ble utarbeidet gjennom å se på eksisterende litteratur på området, hvor det ble identifisert lignende undersøkelser om kartlegging av en data ved en forskningsinstitusjon, og studier om holdninger og motivasjonsfaktorer for å dele forskningsdata.

Tidligere undersøkelser i Norge har i stor grad vært basert på tilpasninger av Parse Insight undersøkelsen (Kuipers et al., 2009), slik det finnes hos Norges forskningsråd (2014a) og L. H. Kvale (2012). Hensikten med disse undersøkelsene har vært å se på forskeres praksis og holdninger til deling av forskningsdata. I Storbritannia har det i større grad vært lagt vekt på å kartlegge en forskningsinstitusjons databeholdning for å sørge for bevaring av forskningsdata, slik som universitetene som inngikk i Data Asset Framework prosjektet (Jones et al., 2009). Denne undersøkelsen har konferert de nevnte studiene og gjort tilpasninger til norske forhold og institusjonelle forhold.

Undersøkelsen legger en litt annen vekt på spørsmålene enn skjema og intervjuer utført ved Georgia Tech (Parham, 2010; Parham, Bodnar, & Fuchs, 2012; Rolando et al., 2013) og i DAF-studiene (Data Asset Framework, 2009; McGowan & Gibbs, 2009). Arbeidet med krav fra universiteter og høyskoler og forskningsfinansiører om at forskere skal utarbeide datahåndteringsplaner er ikke kommet langt i Norge. Foreløpig er det lite sannsynlig at mange forskere har mye erfaring med å utarbeide planer. Forskningsinstitusjonene vil derfor ha liten eller ingen erfaring med dette arbeidet.

Dette har resultert i et spørreskjema med spesielt innspill fra en videreutvikling av DAF undersøkelsene foretatt ved Georgia Tech (Parham et al., 2012) og Norges forskningsråd (2014a). Spørreskjema som anvendes på BI er bestående av følgende tema fordelt over fasene i forskningsprosessen:

1. Den første delen er temaet oppstart av et forskningsprosjekt og finansiering.
2. Den andre delen er handler om arbeidsflyten når data er samles inn og data dokumenteres.
3. Det tredje temaet er knyttet til hva som skjer med data etter avsluttet prosjekt.

4. Den fjerde delen omhandler behov for støtte fra forskningsinstitusjonen ved tilrettelegging av data.
5. Den femte delen er deskriptive data om forskeren som responderer på undersøkelsen.
6. Til slutt inviteres forskeren til å kommentere tema i undersøkelsen eller gi kommentarer til undersøkelsen for øvrig.

Skjemaet finnes i sin helhet i appendiks.

Spørsmålsformuleringene består av åpne og lukkede svaralternativer. Undersøkelsen består for det meste av lukkede svaralternativer. Imidlertid, så vil de fleste av de lukkede svaralternativene har mulighet for å gi skriftlig tilleggsinformasjon. Mange av spørsmålene gir muligheter til å velge flere svaralternativer på et spørsmål. Det er avslutningsvis et åpent svaralternativ hvor respondentene kan gi tilbakemelding på undersøkelsen eller komme med andre bemerkninger.

Det er estimert at undersøkelsen tar ca. 10-15 minutter å besvare. Forskerne som inviteres til å delta i undersøkelsen får tilsendt undersøkelsen på e-post, og vil bli bedt om å utfylle et elektronisk spørreskjema. Undersøkelsen inneholder maksimalt 22 spørsmål, og vises i sin helhet i appendiks. Det nettbaserte verktøyet Qualtrics (2015) ble brukt for å utarbeide spørreskjemaundersøkelsen.

Pilottest av undersøkelsen

Pilottesting kan være med på å klargjøre om funksjonaliteten på nett fungerer som den skal. Brukervennligheten av undersøkelsen testes for å se om respondentene klarer å svare, uten å ha behov for veiledning. Testen er med på å undersøke om ordbruken forstås, eller om det er bruk for oppklarende informasjon. Spørreskjemaet ble pilottestet ut før det ble sendt til forskerne ved Handelshøyskolen BI. Tilbakemeldinger på spørreskjema er blitt gjort av 3 doktorgradsstipendiater som befinner seg på andre studiesteder, ved UiO og HiOA. Stipendiater hører til 3 forskjellige fagområder, bibliotekvitenskap/sosiologi, psykologi og informatikk.

Mottakerne av spørreskjemaet under pilottesten ble stilt noen tilleggsspørsmål. Testerne hadde da mulighet til å gi uttrykk for om de mente at noen av spørsmålene ikke var relevante, eller om det var emner eller spørsmål i undersøkelsen som manglet. Andre kommentarer med innspill til skjemaet var også mulig å gi. Tilbakemeldingene resulterte i at spørreskjemaet ble endret. Det ble fjernet et spørsmål angående tilgang til rådata, endring av spørsmål om

datainnsamling som danner analysegrunnlag, svaralternativer for forskerens ønskede endringer på BI ble redigert slik at det fulgte en prosesstanke, i tillegg ble det lagt til et svaralternativ angående å få «assistance finding av accessing data resources». Noen få skrivefeil ble også rettet.

Kvalitet i undersøkelsen

I vitenskapelige undersøkelser blir kvaliteten bedømt etter validitet og reliabilitet. Validitet er knyttet til i hvilken grad spørsmålene dekker det fenomenet som vi forsøker å undersøke (Østbye et al., 2013). Spørreskjema er ofte velegnet til å kartlegge faktiske forhold. Folks meninger er mer ustabile, slik at svarene blir lettere påvirket av situasjonene, blant annet hvordan spørsmålet stilles. Det henvises til hvordan spørsmålsformuleringen er utformet med hensyn til nøyaktige formuleringer, og om respondentene vil svare oppriktig på spørsmål. Definisjonsmessig validitet i hvor stor grad man klarer å fange overgang fra teoretiske begreper til begrep som brukes i undersøkelsen, dvs. operasjonalisering. Reliabilitet er pålitelighet med hensyn til innsamling, bearbeiding og analyse av data. Sammen utgjør disse to aspektene undersøkelsens validitet (Ringdal, 2013).

Spørsmålene i undersøkelsen er fordelt over kategoriene bakgrunnsspørsmål, atferdsspørsmål og meninger. I undersøkelsen stilles det bakgrunnsspørsmål, slik som respondentenes stillingskategori og avdelingstilhørighet. Dette regnes for å kunne være reliable, da det er sannsynlig at respondentene vil svare troverdig på disse spørsmålene.

Atferdsspørsmål er knyttet til å kartlegge respondentens handlinger, eller ting som han har observert. Undersøkelsen spør om forskningsdata som er samlet inn for et pågående prosjekt eller et nylig avsluttet prosjekt, slik at man kan regne med at respondentene vil ha få problemer med å huske praksis. Av denne grunn tenkes det at svarene vil være reliable.

De fleste faglige ansatte vil være tilknyttet et forskningsprosjekt eller en forskningsgruppe. En del av meningene som etterspørres er relatert til praksis på eget fagfelt, av denne grunn vil det antas at respondentene vil ha en mening omkring temaet. På denne måten vil de skulle tyde på valide og reliable svar. På den andre siden, så er visse aspekter slik som deling av data kommet sterkere i fokus i løpet av den siste tiden. Det kan antas at forskerne kan ha liten personlig erfaring med deling av data slik at forutsetninger for å svare på dette kan være noe begrenset.

Det er ofte vanskelig å få en høy svarprosent ved spørreskjemaundersøkelser. Det finnes, i følge Fink (2013), flere måter å øke responsraten på. Dette kan være å forsikre seg om at

spørsmålene er nøyaktig formulert og forståelig for respondentene. Resultatene må oppbevares konfidensielt og anonymiseres, og man må fortelle hvordan resultatene skal brukes. Tilslutt anbefaler hun at spørreundersøkelser kun anvendes når man tror respondentene er interessert i temaet.

Et annet tiltak som er med på å bidra til høyere svarprosent og øke indre validitet til undersøkelsen er å foreta pilottesting av undersøkelsen. Pilottesting ble som nevnt gjort her.

Utfordringer med nettbaserte undersøkelser er at bortfallet av mulige respondenter kan være stor, slik at svarprosenten blir lav eller skjevhet i utvalget oppstår. Frafallsproblematikken kan reduseres ved å sende ut skjemaer i særskilte eposter, slik at man kan ha en bedre oversikt over hvem som svarer og purre dem som ikke svarer. I undersøkelsen ble respondentens navn brukt på invitasjon til deltakelse for å gi et sterkere personlig fokus. På denne måten vil man kunne identifisere om hvor relevant respondentene synes at undersøkelsen er, eller om det er skjevheter i utvalget. Dette ble det ikke anledning til i denne undersøkelsen på grunn av lister med personlige karakteristika koblet til avdelingstilhørighet ikke var mulig å få fra personalavdelingen.

Generalisering

Undersøkelsesdesignet ved deskriptive tverrsnittstudier har noen begrensninger. Ved at undersøkelsen gjøres på et lite utvalg i en bestemt kontekst på et avgrenset tidsperiode, vil dette gjøre det vanskelig å trekke generaliseringer av resultatet til den totale populasjonen av forskere.

I resultatene kan det by på utfordringer å si om man har funnet en årsakssammenheng, da sammenhenger kan skyldes bakenforliggende variabler. Designet gjør det vanskelig med å si noe om det vil forekomme en endring blant forskeres praksis og meninger over tid, med mindre det gjøres en oppfølgingsstudie.

Etiske betraktninger

Målet med undersøkelsen er å få innblikk i forskerens behandling av forskningsdata. Behandling av data inngår som en del av arbeidsoppgavene til den enkelte forsker.

Behandling av persondata er et aspekt som er aktuelt å se litt nærmere på. I forskningsprosjektet foretas det en innsamling og analyse av data. Det er mulig å koble sammen respondent og arbeidsplass. Siden utvalget er av en liten størrelse, og undersøkelsen foregår innenfor en avgrenset organisasjon, betyr dette at muligheter for å spore tilbake til den enkelte ansatte kan være tilstede.

Det spørres ikke om sensitive opplysninger omkring den enkelte forsker, kun om stilling og avdelingstilhørighet. Bruk av elektronisk spørreskjema gir opplysninger om IP-adresse, slik at dette kan være en identifiserende faktor. Dette behandles med anonymisering av data. I analysen vil heller ikke den enkelte forfatter/ eller informasjon om deres data fremheves, og kan således oppfattes som anonymiserte.

Forskere kan være forsiktig med å uttale seg omkring sine data. Dette gjelder spesielt før resultater er publisert, eller data i selv er personsensitive. Dette kan gjøre at noen forskere ikke ønsker å uttale seg omkring sitt eget prosjekt (Data Asset Framework, 2009). Dette gjelder spesielt ved kvalitative undersøkelser. I denne spørreundersøkelse er dette et mindre problem, da det ikke spørres om innhold eller resultater fra forskernes arbeid. Fokus er på nøytrale kategorier for gruppering av data. Informasjonen som fremkommer under undersøkelsen vil ikke regnes som sensitivt for den enkelte forsker.

Prosjektet er meldt til personvernombudet for forskning⁸. Dette gjøres for at en utenforstående part er med på å sørge for at forpliktelser overfor forskningspersonene er ivarettatt.

Gjennomføring av undersøkelsen

Anvendelse av nettbasert spørreskjema utløser meldeplikt, dersom det er mulig å samle inn personidentifiserende opplysninger gjennom sporing av e-post adresser og IP-adresse. I og med at forskere som er knyttet til en bestemt arbeidsplass gjør identifikasjon potensielt lettere. Dette er også med å utløse meldeplikt. I forkant av undersøkelsen er det søkt personvernombudet for forskning, og laget informasjonsskriv til respondentene med informasjon om konfidensialitet, anonymitet og innhold i skjema, se appendiks B.

Utsendelsene av invitasjon til deltakelse og lenke til undersøkelsen ble gjort via e-postliste til alle faglige ansatte på Handelshøyskolen BI. Spørreundersøkelsen var tilgjengelig for

⁸ Personvernombudet for forskning: <http://www.nsd.uib.no/personvern/om/brukere.jsp>

respondentene i perioden 20.04.15-16.05.15 via det elektroniske verktøyet Qualtrics (2015). Førstegangs utsendelse ble gjort 20.04. Deretter ble det gjort 2 purringer, 03.05 og 13.05. Utsendelsen gav ingen feilmeldinger på e-postadressene. Undersøkelsen ble lukket 16.05. I tillegg ble informasjonsskjermer og intranett brukt til å opplyse om undersøkelsen.

Undersøkelsen er bygget opp omkring følgende tema prosjektoppstart og finansiering, beskrivelse av data knyttet til arbeidsflyten under arbeid med data, databehandling etter avsluttet prosjekt, behov for støtte ved arbeid med data og bakgrunnsinformasjon om stilling og avdeling, og generelle kommentarer om temaet eller undersøkelsen.

Litteraturgjennomgangen legger til grunn en rekke antakelser som belyses gjennom problemstillingene. Forskningen bevares for etterprøving eller nye analyser. Det finnes flere hindre til deling av forskningsdata, hvor årsakene kan være av teknisk, økonomisk, sosial eller juridisk karakter. Erfaring virker inn på holdning til deling av data.

To hovedproblemstillinger ligger til grunn for kartleggingen. Problemstillingene er operasjonalisert gjennom en rekke mer detaljerte spørsmål som henviser tilbake til litteraturgjennomgangen.

- Hvilke forskningsdata finnes på Handelshøyskolen BI?
 - Hvilke filformater og datatyper brukes i organisasjonen?
 - Hvordan blir data lagret?
 - Hvilken praksis finnes det for å dele data?
 - Hvilken praksis finnes det blant forskerne for å gjenbruke andres data?
- Hvordan kan Handelshøyskolen BI som institusjon bidra til forskernes arbeid med data?
 - Hvilke tiltak for forskningsdata mener forskerne BI bør initiere?
 - Hvilke tiltak for forskningsdata mener forskerne er mest aktuelt for seg selv?

Resultater

Dette kapittelet viser funn basert på undersøkelsen, som ble sendt ut på e-post til alle vitenskapelige ansatte ved BI. Resultatene presenteres først med en beskrivelse av respondentene, deres finansieringsløsninger, praksis for databruk og generering av data, praksis for datalagring, hindre for datalagring og forbedringsønsker.

Spørreundersøkelsen ble besvart av 59 av 325 forskere med tilknytning til Handelshøyskolen BI. Dette gav en svarprosent på 18,15 %. I spørreskjemaverktøyet var innstillingen satt til «force response». Dette innebar at respondentene må besvare et spørsmål for å gå videre til neste spørsmål. Påbegynte spørreskjema, som ikke er fullført, er ikke inkludert i undersøkelsen. Dette utgjorde 21 påbegynte skjema, hvor 7 stykker kun så på innledningen til skjemaet. En respons manglet besvarelse på det siste spørsmålet omkring stilling, er inkludert blant de 59 svarene.

Respondentene fordelte seg over alle fagavdelingene ved Handelshøyskolen BI, som vist i tabell 1. Flest svar kom fra respondenter ved avdelingen «Strategy and logistics» (12 svar), mens færrest svar kom fra «Accounting, auditing and law» (5 svar) og «Communication and Culture» (5 svar).

Tabell 1. Respons basert på avdelingsvis fordeling

	Inviterte / populasjon	Svar	Responsrate
Strategy and logistics	36,7	12	20%
Marketing	29	10	17%
Leadership and organizational behaviour	47	9	15%
Economics	38,9	8	14%
Finance	31,2	6	10%
Accounting, auditing and law	51,3	5	8%
Communication and culture	24,1	5	8%
Innovation and economic organization	37,1	4	7%
Total	295,3	59	100%

Note. N = 59. Inviterte/populasjon basert på tall fra DBH.

Respondentenes stillingskategori og avdelingstilknytning er presentert i tabell 2.

Tabell 2. Respondentenes stillingskategori og avdelingstilknytning

	Accounting, auditing and law	Communication and culture	Economics	Finance	Innovation and economic organization	Leadership and organizational behaviour	Marketing	Strategy and logistics	Total
Professor	2	2	3	3	1	5	2	3	21
Associate Professor	1	1	-	1	2	2	6	5	18
Researcher	-	-	-	-	-	-	-	-	-
Senior Lecturer	-	-	-	-	-	-	1	-	1
Lecturer	-	-	-	-	-	-	-	-	-
Postdoctoral Fellow	2	1	1	2	-	-	-	-	6
Ph.D Student	-	-	4	-	1	1	1	2	9
Other	-	1	-	-	-	-	-	2	3
Total	5	5	8	6	4	8	10	12	58

Note. N = 58. Inndelingen i stillingskategorier er basert på ansatte kategorier som har vært anvendt i undersøkelser i norsk høyere utdanning tidligere (thune).

Respondentene har i stor grad avansert lenger opp i universitets- og høyskolesystemet. De fleste som har svart på undersøkelsen innehar stillinger som førsteamanuensis (18 svar) eller professorer (21 svar), slik at dette kan indikerer lengre erfaring (til sammen 39 svar). Mens noe kortere erfaring finnes hos øvrige respondenter (15 svar) som er postdoc (6 svar) eller phd-kandidater (9 svar).

Den første delen av spørreskjema er knyttet til aktiviteter som foregår før eller ved oppstart av et forskningsprosjekt. Fokus på finansiering ved prosjektoppstart er interessant fordi krav fra EU og Forskningsrådet sier at forskere skal ha en datahåndteringsplan, hvor krav om tilgjengeliggjøring av data er knyttet til materiale som forskeren selv samler inn. Det første spørsmålet omhandler om forskerne har finansiering for undersøkelser/datainnsamlinger som de selv gjør. På dette «ja» eller «nei»- spørsmålet viste det at blant de 59 forskerne (100 %) som har svart på spørreskjemaet, er det 28 (47 %) som har forskningen sin finansiert på en eller annen måte, mens 31 (53 %) har egenfinansiert forskning. De har forskningen finansiert ble bedt om å spesifisere kilde i et tilleggsspørsmål. Variablene i spørsmålet anvender inndeling i finansieringskilder som forskere i Norge ofte får støtte fra, lignende inndeling kan finnes i Thune, Aamodt, og Gulbrandsen (2014). Svar på spørsmål 1a «Who is the funding agency/source?» er vist i tabell 3.

Tabell 3. Finansieringskilder for pågående prosjekter blant respondentene

Variabel	Respons	
	n	%
Norwegian Research Council	11	39%
Governmental/ public agencies	4	14%
Industry/ business	5	18%
Private foundation / organisation	0	0%
BI Norwegian Business School	7	25%
EU programs	2	7%
Other Norwegian Source	1	4%
Other International Source	3	11%
Total	33	118

Note. N= 28 Annen norsk kilde = CCGR,

Annen internasjonal kilde: US National Institute of mental health, Project management institute

Det er mulig å krysse av for flere svar på dette spørsmålet, slik at dette reflekterer at forskerne (28 svar) i noen tilfeller (5 svar) får finansiering fra flere kilder (33 finansiører). Norges forskningsråd er den største aktøren hvor forskerne får sin forskning finansiert 11 prosjekter (39 %). Deretter får 7 av respondentene sine prosjektmidler tildelt av BI (25 %). Det er ingen av respondentene som får midler fra private stiftelser. Som pekt på tidligere, stiller forskningsfinansiørene ofte krav ved søknad om midler om datahåndteringsplan. Dette gjelder Forskningsrådet, EU og øvrige internasjonale kilder. Dersom respondentene skulle søkt midler i dag vil 16 av 33 (20,6 %) hatt krav om utforming av DMP.

Finansiering av forskningsprosjektet kan ha betydning for om eierskap til data og muligheter for pålegge data lisensbestemmelser. Forskere med finansiering fra NFR og EU pålegges krav om lagring for data. Finansieringsspørsmålene følges opp med et spørsmål omkring hvordan data samles inn. Spørsmål 2. «What origin has data used as basis for your research analysis?». Variablene er utledet fra inndelingen av data fra National Science Board (2005). På denne måten ledes respondenten over til den andre delen av spørreskjemaet omhandler beskrivelse av data som benyttes.

Tabell 4. Datatyper basert på innsamlingsmetode

Variabel	Respons	
	n	%
My data are observational data, i.e. interviews, survey, instrument data	34	58%
My data are register data, i.e longitudinal public, health data	19	32%
My data are records, i.e. documents, movies	10	17%
My data are experimental data, i.e resulting from experimental procedures	8	14%
My data are computational data, i.e. computer modelling, simulations	6	10%
Other	5	8%
Total	82	139%

Note. N=59.

De fleste forskerne (58 %) på BI produserer observasjonsdata. Deretter er registerdata (32 %) det mest brukte grunnlaget for analyser og studier. Bruken av flervalgsalternativet ved svar på spørsmålet viser at mange forskere kombinerer forskjellige datainnsamlingsmetoder, til bruk for metodetriangleringer eller bruk av andres data i kombinasjon med egne data.

Observasjonsdata vil forskeren/forskningsgruppen ha råderett over, mens registerdata kan ha egne lovmessige regler og begrensninger knyttet til seg. I og med at digitale data som innsamlet som primærdata er fokus for forskningsfinansiørene kan kombinasjoner av data fra flere kilder skape diskusjoner omkring rettigheter og eierskap til data med hensyn til krav om arkivering når forskeren skal utøve praksis.

Innsamlingsmetoden er bare et av elementene som tilslutt kan påvirke muligheten for lagring av forskningsdata, slik som det er satt av forskningsfinansiørene. For å kunne arkivere forskningsdata er dette knyttet til filformatene som benyttes. Finansiørene peker på viktigheten av å arkivere data. Filformater som er lite brukt vil ofte ha utfordringer med å kunne finne et arkiv som kan lagre og oppdatere filformatene over lengre tid. I spørsmål 3. "Choose all of the following formats that best describe your research data" blir respondentene bedt om å beskrive filformatene som de anvender. Det er gitt noen eksempler på filformater gjennom forkortelsene ved lagring av dokumenter, for eksempel er .doc en henvisning til tekstdokumenter generert i Microsoft Word. Svarene er vist i tabell 5.

Tabell 5. Filformater som genereres av forskerne

Variabel	Respons	
	n	%
Numerical scores (.xls, .csv, .sav, .R)	50	85%
Textual records (.doc, .rtf, .txt)	19	32%
Databases (.db, .mdb, .sql)	15	25%
Networked based data and web files (.html, .xhtml)	5	8%
Images, sounds, videos and graphics (.jpg, .mov, .avi, .mp4)	5	8%
Scripts and programming (Java,C++)	4	7%
Other (please specify)	2	3%
Geographical information files (.gpx, .kml, .kmz)	0	0%
Total	100	168%

Note. N = 59. Filformater Other er nevnt som SPSS, pajek

Resultatene viser filformatenes bruk blant respondentene. Nummeriske filer (85 %), for eksempel ulike typer regneark, er det som hyppigst genereres av forskerne som beskrivelse av deres data. Etterfulgt av tekstfiler (32 %) og databaser (25 %). Blant respondentene er det få som anvender nettbaserte filer (8 %), billed/lydformater (8 %) og skriptbaserte filer (7 %). Ingen av respondentene bruker geografiske filer. Dette indikerer at filformater som brukes til å generere data er knyttet til de vanligste kontorstøtteprogramvarene og analyseverktøyene innen forskning.

I arbeid med data er filstørrelser av betydning for lagringskapasitet under og etter et forskningsprosjekt. Respondentene ble i spørsmål 4 bedt om å indikere hvor mye data deres prosjekt ville danne. Tabell 6 viser svaret på «Indicate the approximate amount of data the project is expected to generate».

Tabell 6. Estimerte filstørrelse for respondentenes forskningsprosjekter

Variabel	Respons	
	n	%
1 - 500 gigabytes (GB)	32	54%
> 500 - 1000 GB	3	5%
1 - 500 terabytes (TB)	1	2%
> 500 - 1000 TB	0	0%
1 - 500 petabytes (PB)	1	2%
> 500 PB	0	0%
Don't know	18	31%
Other (please specify)	4	7%
Total	59	100%

Note. N = 59. Other: "This is really hard to say, as our present approach is testing the borders of computability and this is likely to increase", "Not sure; just know the data base is big", "1 kb", "119KB"

32 respondentene (54 %) genererer ikke filstørrelser gjennom prosjektene sine som er meget store. Lokal IT-avdeling vil ha behov for å kunne ivareta datamengder for mange prosjekter som er av mindre størrelse. Prosjekter som genererer enda større datamengder (9 %) vil ha behov for et særskilt fokus på filstørrelser gjennom måten forskningen foregår på, slik at dette kan være nødvendig å tilpasse i de enkelte tilfeller. Likevel kan resultatene indikere at per dags dato er dette en utfordring kun et fåtall av prosjektene møter. Det er imidlertid en stor andel (31 %) som ikke vet hvor store filstørrelser prosjektene genererer.

For forskerne kan tap av data på grunn av manglene lagring eller ødeleggelse av filer hindre fremgangen i forskningen. Respondentene har svart på spørsmål 9 «If your research data gets lost, how easily can you recreate it?», vist i tabell 7.

Tabell 7. Vanskeligheter ved å gjenskape data

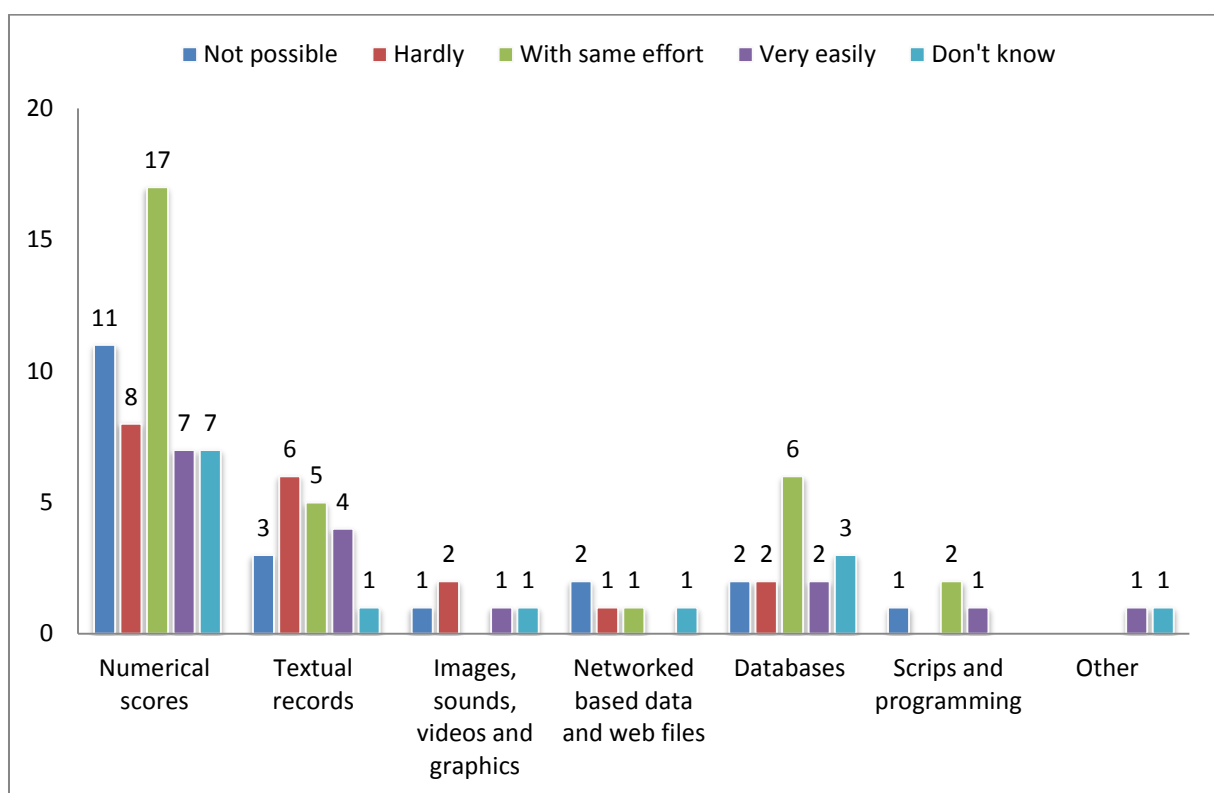
Variabel	Respons	
	n	%
Not possible	13	22%
Hardly	11	19%
With same effort	18	31%
Very easily	10	17%
Don't know	7	12%
Total	59	100%

Den alvorligste konsekvensen er at data ikke kan gjenskapes. 13 av respondentene (22 %) peker på at data ikke er mulig å gjenskape, mens 11 svar sier det er vanskelig å gjenskape (19 %). Dette kan tyde på at sikker lagring av data bør være av spesiell interesse for 24 respondenter (41 %), da det vil ikke være mulig å gjenta forskningen eller innebære betydelig

merarbeid. Likeledes peker 18 respondenter (31 %) på at data kan enkelt gjenskapes med den samme innsatsen eller med mindre innsats. 28 respondenter (48 %) vil kunne gjenskape samme resultater med samme innsats eller mindre innsats. En god del av respondentene (12 %) peker på at de ikke vet hvilke konsekvenser for arbeidsmengde eller forskningen det vil ha å gjenskape data, dersom de tapes.

Det kan diskuteres om bruk av noen filformater er lettere å gjenskape. Fordelingen av filformater og reproduksjon ved tap av data er vist i figur 1.

Figur 1. Filformater og reproduksjon ved tap av data



Bruk numeriske data er filformatet som 11 av respondentene finner vanskeligst å gjenskape, men er også det enkleste filformatet å gjenskape i følge 7 respondenter. Å investere samme innsats for å gjenskape tapte data er gjennomgående beskrevet som arbeidsmengde.

Respondentene ble spurt om de hadde brukt data fra andre som grunnlag for sin forskning i løpet av de siste 3 årene. I løpet av en 3 års periode svarte 30 respondentene (51 %) positivt på at de har brukt andre forskeres data, mens 29 stykker (49 %) har ikke brukt andres data.

Halvparten av respondentene som gjenbraker data ble stilt spørsmål 5a “Whose research data have you used the most within the past 3 years?”. Hvem sine data som gjenbrukes er vist i tabell 8.

Tabell 8. Forskeres bruk av data produsert av andre forskere

Variabel	Respons	
	n	%
Research data from other researchers at my institution	12	40%
Research data from other researchers at national institutions	11	37%
Research data from other researchers at international institutions.	15	50%
Total	38	127%

Note. N = 30.

Blant forskerne som har brukt andres data, er det mest brukt data fra forskeren ved internasjonale forskningsinstitusjoner (50 %), tett etterfulgt av data fra forskere ved BI (40 %) og nasjonale institusjoner (37 %).

Det tredje temaet i spørreundersøkelsen omhandler hva som skjer med data etter avsluttet prosjekt. Ved lagring av data er gjenfinning av data en forutsetning for gjenbruk av data. Gjenbruk forutsetter tilstrekkelig dokumentasjon gjennom beskrivelser, og et klart forhold til hvem som har eierskap til data og lisenser for bruk. Derfor er det viktig å vite noe om datas mulige tilgjengelige lagringsmåter/medium, og eventuell bruk av dataarkiver. Respondentene ble i spørsmål 6 bedt om å besvare “What is the most common way of archiving your research data after results are ready or beyond the life of a project?”

Tabell 9. Dataarkivering etter avsluttet prosjekt benyttet av forskerne

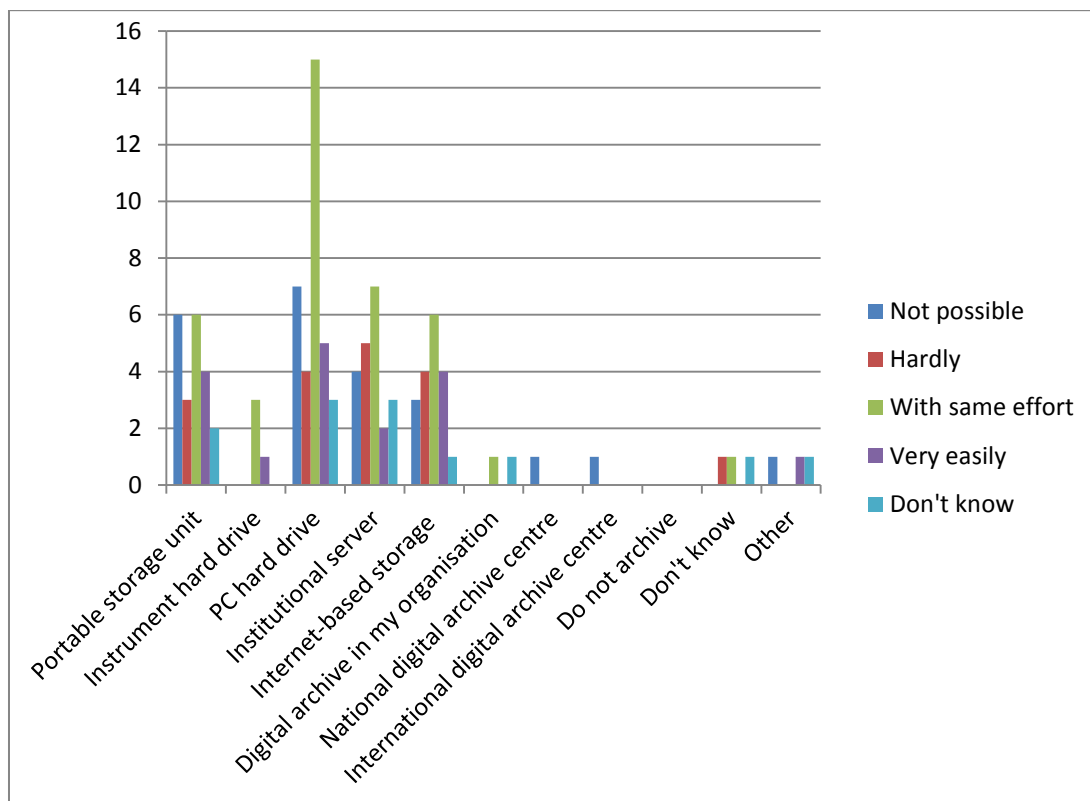
Variabel	Respos	
	n	%
PC hard drive	34	58%
Portable storage unit (i.e., USB flash drives, CD/DVD, External hard drive)	21	36%
Institutional server	21	36%
Internet-based storage (e.g., cloud or grid storage)	18	31%
Hard drive of the instrument which generates the data	4	7%
Data is submitted to digital archive centre in my organisation	2	3%
Data is submitted to a national digital archive centre	1	2%
Data is submitted to an international digital archive centre	1	2%
Don't know	3	5%
Other (please specify)	3	5%
Do not archive	0	0%
Total	108	185%

Note. N = 59. De som bemerker Other spesifiserer dette som «Own web page», «I intend to destroy the data after the project completion», «Journal may require copy of data».

Alle respondentene svarte at de ivaretar lagring av data etter avsluttet prosjekt, vist i tabell 9. Variasjonen er knyttet til hvilket lagringsmedium som anvendes. De fleste forskerne lager data på egen PC (58 %), på bærbare lagringsenheter (36 %), eller på institusjonell server (36 %). Bruk av internettbasert lagring som skytjenester (31 %) er utbredt. Det er nesten ingen av respondentene som lager materiale hos et dataarkiv, institusjonelt, nasjonalt eller internasjonalt (7 %). En liten andel (5 %) er usikker på hvor dataarkivering gjøres. Respondentene har data i stor grad lagret mer enn ett sted. Bruk av flere lagringsenheter kan innebære at muligheten for å kunne gjenfinne data for den enkelte forsker kan være bra, forutsatt det er snakk om at samme versjon av data lagres flere steder. Resultatet impliserer at data i liten grad kan gjenfinnes for andre enn den enkelte forsker.

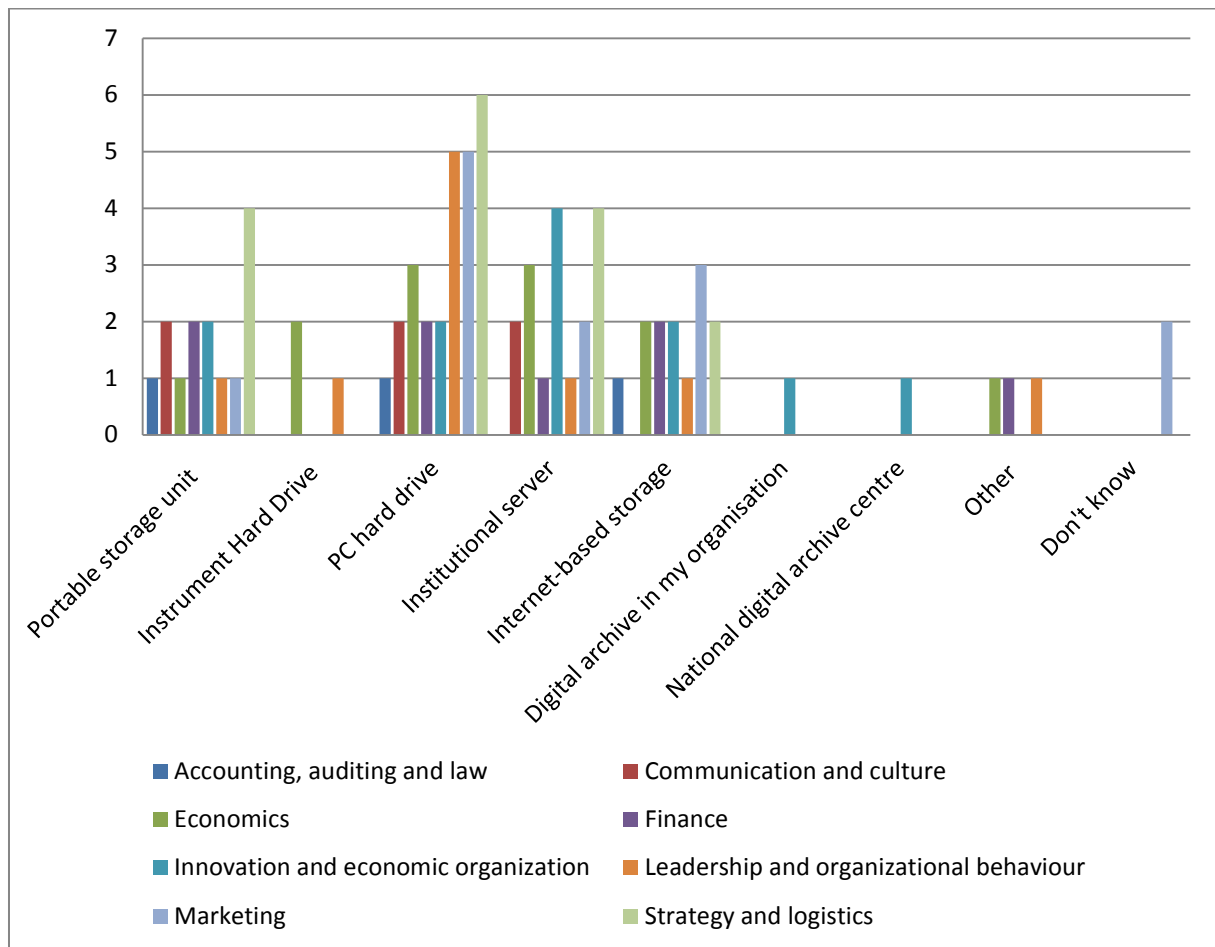
Hvordan respondentene bruker de forskjellige lagringsmediumene er vist i figur 2. Logikken burde tilsi at dersom data er vanskelig å gjenskape, så bør det lagres sikkert. Bruken av dataarkiver er som nevnt liten (4 svar), hvor 2 peker på at arkivering i dataarkiv anvendes når data ikke er mulig å gjenskape. Data som er vanskelig å gjenskape lagres for øvrig på bærbare enheter, PCs harddisk, institusjonell server og via skytjenester. Data som vanskelig kan gjenskapes lagres ikke på instrumentet som har bidratt til datainnsamlingen.

Figur 2. Lagringsmåte for data sett i forhold til mulighet for å gjenskape data dersom data i en undersøkelse forsvinner



Kan lagring av data utøves forskjellig på de ulike avdelingene? Bruk av lagringsenheter fordelt på avdelingstilhørighet, vist i figur 3.

Figur 3. Bruk av lagringsenheter fordelt på avdelingstilhørighet



Note. På grafen vises ikke «Do not archive» og «International data archives» er fjernet, da ingen av respondentene hadde benyttet svaralternativene.

Når man ser på bruken av lagringsenheter fordelt på avdelingstilhørighet, kan man se at med hensyn til lagringsløsninger synes noen avdelinger å skille seg litt fra de øvrige avdelingene. 2 respondenter fra «Innovation and economic organization» er de eneste som bemerker at de benytter dataarkiver. Likeledes er lagring på instrumentet som bidratt til datainnsamling kun anvendt blant respondentene to avdelinger «Economics» og «Leadership and organizational behaviour», selv om dette kun gjelder 3 svar. Bruken av skytjenester er utbredt, men ingen av respondentene ved «Communication and culture» har bemerket at de bruker dette. Skytjenester anbefales generelt ikke fra mange IT avdelinger. Det er usikkert hvem som har tilgang til data og ikke alle skybaserte lagringsløsninger lagrer alle versjoner tilbake i tid, eller gir back-up for data.

Forskernes tilfredshet med eksisterende lagringsløsninger inngår i spørsmål 11: «Er du tilfreds med arkiveringsløsningen som du bruker i dag?» var formulert som et ja-nei spørsmål. 43

respondenter (73 %) var tilfreds med dagens arkiveringsløsning, mens 16 respondenter (27 %) var ikke tilfreds. Årsakene til hvorfor forskerne ikke er tilfreds er spesifisert i tabell 10.

Tabell 10. Manglende tilfredshet med dagens arkiveringsløsning

11a. If “not”, why are you not satisfied with your archiving solution?
(Multiple answers allowed)

Variabel	Respos	
	n	%
Too little capacity	6	38%
Not secure enough	6	38%
Too many archiving solutions	5	31%
Other (please specify)	5	31%
Too complicated to use	4	25%
Too few archiving solutions	1	6%
Too expensive	1	6%
Total	28	206 %

Note. N = 16.

Other = “Chaotic and unsystematic today”, “What archiving - none exists”, “Need both high security and easy access from outside, we chose the former”, “Could be better interface and easy overarching solution”, “Data comes from different sources and most in formats not tailormade for research”. “Have to figure out how to manage and organize the data, and know too little about what is optimal”.

Respondentene som er misfornøyd fremholder at det er mer enn en årsak til deres misnøye. Liten kapasitet (38 %) og manglende sikkerhet (38 %) pekes det på.

Likeledes synes en av utfordringene synes å være mange lagringsløsninger, da 31 % av respondentene synes dette. Samtidig synes få arkiveringsløsninger ikke å være et betydelig problem. Dette kan implisere behov for hjelp til å velge rett løsning for data og deres praksis.

30 (51 %) av respondentene svarte at de har brukt data fra andre som grunnlag for egen forskning. I undersøkelsen fikk respondentene anledning til å peke på hvem de ønsker å dele data med gjennom spørsmål 10 «Identify all of the following who you would like to share anonymized data with, if you had the ability to do so”.

Tabell 11. Deling av anonymiserte data

Variabel	Respons	
	n	%
Researchers, students, and staff working on the project	53	90%
Other members of the affiliated School or Research Center	24	41%
Researchers at other institutions	28	47%
Project sponsors	11	19%
General public	12	20%
Total	128	217%

Note. N=59.

Forskerne viser en stor villighet (90 %) til å dele anonymiserte data til forskere, studenter og andre som arbeider i det samme forskningsprosjektet. Deretter ønsker de å dele med andre forskere ved andre institusjoner (47 %), så andre forskere ved egen institusjon (41 %). De færreste er interessert i å dele data offentlig (20 %) eller med oppdragsgivere (19 %).

I praksis hviler det en rekke utfordringer med å gjøre data tilgjengelig for andre forskere. Dette ble belyst gjennom spørsmål 12 “Do you see any challenges in making more of your research data available for other researchers?”

Tabell 12. utfordringer med å gjøre data tilgjengelig for andre forskere

Variabel	Respos	
	n	%
I cannot give access due to sensitivity issues	28	47%
Preparing data for open access takes away valuable time for research	21	36%
Reduce possibilities of future scientific publications	20	34%
I am afraid other researchers will not understand my data	15	25%
Lack of technical infrastructure	11	19%
I cannot give access due to shared ownership	10	17%
I am afraid data will be misused	8	14%
It would be unethical	8	14%
I cannot give access due to commercial issues	7	12%
I do not believe data is secure at a data centre, journal site or alike	6	10%
I cannot give access due to intellectual property rights	5	8%
Open Access to research data might have a negative economic impact for me and my institution	5	8%
I do not know	5	8%
Other (please specify)	4	7%
I do not believe my research data is of interest to others	2	3%
Total	155	262 %

Note. N = 59. Other = "It would be illegal!", "No", "I don't see any challenges", "The data are confidential data from the Norwegian tax authorities and cannot be shared with others".

I tabell 12 peker respondentene på hvorfor de tenker deling av data er vanskelig. Deling av data er først og fremst vanskelig på grunn av anonymiseringsaspekter (47 %), deretter tidsbruk for tilrettelegging av data (36 %), reduserte muligheter for publisering (34 %). Deretter følger frykt for at andre kan misforstå data (25 %), manglende teknisk infrastruktur for deling (19 %), og felles eierskap til data (17 %).

Dersom man ser nærmere på data henviser 59 svar (100 %) til etiske aspekter. Dette synes å være den største hindring for deling av data. Sett med hensyn til variablene sensitivity (47 %), misunderstanding (25 %), misuse (14 %), unethical (14 %). Etikk vektlegges mer enn rettigheter til data 22 svar (37 %) som en hindring for deling. Rettigheter summert av variablene shared ownership (17 %), commercial issues (12 %) og intellectual property rights (8 %).

Flere vitenskapelige tidsskrifter stiller krav om tilgjengelighet av data som en del av deres publikasjonsprosess. Spørsmål 13 ble stilt som "Do you welcome the trend of making data available as a part of scientific publications?" Respondentene svarte på deres syn på tilgjengeliggjøring av data som del av vitenskapelige publikasjoner, som vist i tabell 13.

Tabell 13. Forskernes syn på data som del av vitenskapelige publikasjoner

Variabel	Respons	
	n	%
Yes, it could mean that my research could be more interesting for others to follow	22	37%
Yes, it is a sign that my research can be quality assured	15	25%
No, I see no benefit for me	12	20%
I do not know	5	8%
Yes, it could mean that my data and or my publications will be more cited	3	5%
Other (please specify)	2	3%
Total	59	100%

Note. N = 59.

Other = "Maybe it is possible to map other data expelling other aspects", "More transparency as well as easier access to data for meta studies".

Forlag/tidsskrifters krav om publisering av data sees på i et perspektiv hvor egen forskning blir mer interessant for andre å følge (37 %), og en faktor som kan være med på å kvalitetssikre forskningen (25 %). De positive aspektene (62 %) for den enkelte forskerens posisjonering og faglige kvalitet er sterkere enn andelen forskere (20 %) som mener at tidsskrifters krav om dataarkivering har ingen fordeler. Det er liten tro på at publisering av data er en medvirkende faktor til at deres data eller publikasjoner vil bli mer sitert (5 %).

Innen arkivsektoren regnes lagring med varighet over 10 år som langtidslagring. Ved lagring over et slikt tidsrom vil tilpasning til ny programvare og filformater oppstå. I spørsmål 7 ble respondentene bedt om å spesifisere hvor lenge de ønsket å bevare data, vist i tabell 14. 21 respondenter (35 %) mener lagring over ti år, dvs. langtidslagring, er ønskelig. 34 respondenter (65 %) mener at det kan være mer hensiktsmessig med korttidslagring. Dette indikerer at respondentenes egne vurderinger for lagring over tid er mindre enn kontraktskrav fra Forskningsrådet.

Tabell 14. Respondentenes beregning av hvor lenge data bør lagres

Variabel	Respons	
	n	%
Less than 3 years	7	12 %
3 - 6 years	17	29 %
7 - 10 years	14	24 %
11 - 20 years	9	15 %
More than 20 years	12	20 %
Total	59	100 %

Note. N= 59.

Deretter i spørsmål 8 blir det spurt om årsakene til hvorfor respondentene ønsker å bevare data. Svarene fra fritekstspørsmålet omfavner flere aspekter ved lagring fra forskerens standpunkt. Forskerne tar vare på data først og fremst for å kunne dokumentere disse, med tanke på etterprøvnbarhet og gjenbruk i nye analyser. Gjenbruk er knyttet til å bruke data på ny, for eksempel gjennom å gjøre analyser i kombinasjon med andre data eller nyere data. Men gjenbruk med tanke på at andre forskere kan bygge videre på data nevnes.

Etterprøvnbarhet av data vektlegges stort. Dette gjelder muligheten til å replisere egne data eller ha data tilgjengelig dersom andre stiller spørsmål omkring data slik at man må «defend findings». Publisering er en sentral årsak til å beholde data. Data må oppbevares lenge nok til å kunne gjøre få gjort første publisering av artikler basert på data. Tidsskrifters krav hvor det etterspørres data er også blant årsakene til å beholde data.

Det bemerkes i flere av svarene at det ikke er interessant å beholde data lenge i alle tilfeller, eller at data vil «utgå» på dato, «use again too old». Mens andre mener at data bør beholdes til evig tid «data should in principle be kept indefinitely». Tabell 15 viser hyppige argumenter for behov for lagring og bevaring av data. Dette er et utvalg av alle kommentarene, en fullstendig oversikt finnes i appendiks.

Tabell 15. Hyppige argumenter for behov for lagring og bevaring av data

Fremsatte argument	
Gjenbruk av data	<p>In case further interesting analyses can be made, e.g., if the data could be connected with other data sets in the future.</p> <p>Raw data can be useful for other researchers or for myself in my future researches. It is always good to have (as much as possible) not-biased raw data (i.e. data that can be interpreted in another way, on the basis of a different theory, a different point of view, etc.)</p> <p>Valuable data which was gathered also manually from records, can be used for new projects</p> <p>The data I use is all publicly available. The main reason for me to keep it myself is in case I want to reuse it in future projects. The total size is also quite small, so keeping it is basically costless.</p>
Etterprøvbarehet	<p>There is an increased interest in sharing data, asking for renewed submissions in case of doubts and protests. All practices indicate that researchers will be increasingly accountable for documenting the veracity of their data in the years to come.</p> <p>Allow other researcher to replicate and build on my research. Also easy for myself to reproduce my own research.</p> <p>Until interest in the work has subsided, any disagreement has surfaced, and I would no longer be required to defend findings.</p> <p>For research integrity, the data should be available for other researchers/academic journal; and for potential future project</p> <p>In case journal asks for a review of results.</p>
Lagringstid	<p>Forefront research should be actual for many years</p> <p>Data should, in principle, be kept indefinitely.</p> <p>Contracts with Statistics Norway and Centre of Corporate Governance Research (CCGR) at BI specify (tentative) dates of deletion of research data, often agreed on from our expected publication dates.</p> <p>Sometimes it is not allowed to keep the data longer. But I would be happy to keep the data for longer time since one can have many ideas by using the same data and it takes time from starting the papers to getting them published.</p>

Den fjerde delen av undersøkelsen omhandler behov for støtte ved tilrettelegging av data. Å tilrettelegge data og gjennomføre analyser krever mange kunnskaper, slik at det er av interesse å vite noe omkring forskernes egen interesse for å lære mer og få støtte fra andre i denne prosessen. Spørsmål 14. «What effects do you think BI should initiate to improve preservation, curation and sharing research data?» viser til forskernes ønsker om tiltak ved BI, se tabell 16.

Tabell 16. Forskernes ønsker om tiltak ved BI

Variabel	Respons	
	n	%
Better infrastructure for research data	26	44%
More resources allocated to research data activities	21	36%
More training on the topic research data	15	25%
I do not know	15	25%
Other (please specify)	12	20%
Implementation of guidelines	9	15%
Implementation of standards	6	10%
Make open access to research data an indicator in the funding scheme	5	8%
Implementation of a system for citation	3	5%
Total	112	188%

Note. N= 59.

Other:

We should have experts on database management that could help us.

Funding for compiling data at the first place

Get an institution contract with SSB like Frisch center!

Do not have time to set this up

Provide better access to databases

Encourage journals to make data for published articles available

My qualitative data should not be available for others due to anonymity issues

I don't think it is necessary to do anything. If any researcher is interested in my data, they should just ask for it via email.

I think no action is required. People know how to take care of data and if be asked to provide their data, they will to the specific individuals who have asked for it. No need to publicly sharing!

BI should not make any effort

No efforts needed

None should be used. This is handled by the journals after publication. There exist international platforms that we can use, such as www.runmycode.org. Journals like Marketing Science and International Journal of Marketing Research encourages its authors to publish data and syntax there. BI might ask faculty at BI to post our data there. An issue I have not considered, however, is the issue of scientific fraud. Some schools require faculty to store datasets on the schools servers/systems as a way to avoid/control fraudulent practices.

Bedre infrastruktur (44 %), flere ressurser til arbeid med forskningsdata (36 %), mer opplæring (25 %) peker respondenter på at det er ønskelig at BI initierer.

En andel av respondentene, 15 svar (25 %) er usikker på hva BI bør gjøre med forskningsdata. I feltet «other» kommer det frem 5 kommentarer som peker på at det ikke er nødvendig at BI gjør noe med forskningsdata. Svaret på dette spørsmålet gav en utstrakt bruk av spesifisering av svaralternativet Other. Kommentarene knytter seg til ønske om en politikk ved BI med ingen tiltak, samtidig som det ønskes å legge press på tidsskrifter for at de skal føre en politikk for tilgjengeliggjøring av data. Noen av kommentarene er knyttet til infrastruktur. Dette kan være valg av arkiv eller bedre tilgang til databaser med et datagrunnlag som kan

brukes i forskning. Arbeidssituasjonen for forskere er henvist til gjennom et ønske om flere formelle reguleringer og støttefunksjoner men hensyn til utførelse av analyser.

Videre er det sett på forholdet mellom ønskede tiltak på BI i forhold til avdelingstilhørighet i tabell 17.

Tabell 17. Ønske om tiltak for BI fordelt over forskernes avdelingstilhørighet

	Accounting, auditing and law	Communication and culture	Economics	Finance	Innovation and economic organization	Leadership and organizational behaviour	Marketing	Strategy and logistics	Total
Better infrastructure for research data	3	2	4	1	3	3	4	6	26
Implementation of a system for citation	1	0	0	0	0	1	0	1	3
More resources allocated to research data activities	3	1	4	2	0	4	2	5	21
Implementation of guidelines	0	2	1	0	1	2	2	1	9
More training on the topic research data	1	1	2	0	2	3	2	4	15
Implementation of standards	0	1	2	0	0	2	0	1	6
Make open access to research data an indicator in the funding scheme	0	1	2	0	0	1	0	1	5
I do not know	1	2	2	1	0	3	4	2	15
Other (please specify)	1	0	1	3	1	1	3	2	12
Total	5	5	8	6	4	9	10	12	59

26 respondenter over alle avdelinger ønsker seg bedre infrastruktur for arbeid med forskningsdata. Det er en bred enighet over alle avdelinger 21 svar med unntak Innovation and economic organization, om at det er ønskelig med mer ressurser til forskningsdataarbeid. Opplæring er ønsket av 15 respondenter ved samtlige avdelinger, unntatt Finance.

De øvrige tiltakene har respondentene mindre fokus på. 9 svar nevner behov for retningslinjer fordelt på respondenter fra alle avdelinger, unntatt Economics og Finance. Standarder ønsket av 6 respondenter fra Communication, Economics, Leadership and organizational behaviour og Strategy and logistics. Mens implementasjon av siteringssystem, er kun bemerket behov fra 1 svar fra henholdsvis Accounting auditing and law, Leadership and organizational behaviour og Strategy and logistics. Usikkerheten omkring hva som kan være tiltak for BI, er jevnt fordelt ut over alle avdelinger med unntak av innovasjon.

Tabell 18 legger mer vekt på forskerens ønsker om tiltak for egen utvikling.

Tabell 18. Forskernes ønsker om tiltak for egen utvikling

Variabel	Respons	
	n	%
Data storage and backup while research is on-going	26	44 %
Assistance finding and accessing data resources	20	34 %
Information regarding data management best practices	19	32 %
Assistance to meet data sharing or data management requirements of funding agencies	18	31 %
Provide long-term data access and preservation	18	31 %
Assistance using tools for information gathering and analysis	14	24 %
Tools for sharing research data as research is on-going	14	24 %
Information about developing a data management plan	12	20 %
None of the above	9	15 %
Assistance selecting data to preserve for the long-term	8	14 %
Assistance applying metadata to research data	6	10 %
Information about citing data resources	4	7 %
Other (please specify)	3	5 %
Total	171	291 %

Note. N=59.

Other: "No services needed", "Information on security", "BI should not do this"

Spørsmål 15 «Identify which of the following services might be useful in regards to the management of research data». Tiltak for egen utvikling innen feltet forskningsdata har fokus på tiltak ved pågående forskning, dvs. før prosjektavslutning. Løsninger for lagring og back-up (44 %), assistanse på å finne frem og få tilgang til dataressurser (34 %) og «best practice» for RDM (32 %). Det er noe mindre fokus på hva man skal gjøre med data etter forskningsprosjektet. Dette kan være slik som gi bevaring og lagring på data over tid (32 %), utvalg av data for langtidslagring (14 %) og hjelp til bruk av metadata (10 %).

Undersøkelsens siste spørsmål er en mulighet til å gi åpne kommentarer om administrasjon, lagring av data eller generelt om undersøkelsen. De fleste kommentarene er generelle kommentarer omkring forskningsdata og hva som forskerne er opptatt av at BI bør bistå med. Spredningen på kommentarer tyder på at hele spekteret representeres i kommentarene. Fra forskere som mener at BI ikke behøver å gjøre noe med forskningsdata og at dette er forskerens eget anliggende til dem som synes dette er et viktig tema: «No initiation is really necessary.» “All good researchers are already taking care of their data. If any journal or other researcher needs our data, they can just email us.» Flere kommentarer er knyttet til tilgang på data. Det er sen og kostbar prosess for å få tilgang til registerdata, og kjøpe flere tilganger til databaser som inneholder informasjon som kan brukes som grunnlag for analyser. Muligheter til å få tilgang til forskningsobjekter som kan legge grunnlag for observasjonsdata er en annen frustrasjonsfaktor. Øvrige kommentarer er knyttet til muligheter for opplæring innen

Research Data Management og støttefunksjoner for statistiske analyser. Alle kommentarer finnes i appendiks.

Med hensyn til undersøkelsen kommenteres et spørsmål om anonymitet, da stillingskategoriene er fininndelt, og ved noen av instituttene er det ikke så mange stillinger innen hvert område. Det ble foreslått å ta med mer bakgrunnsspørsmål, slik som stillingsprosent. Undersøkelsen gikk ut til alle faglige ansatte, slik de som mottok undersøkelsen inkluderer personer i undervisningsfokuserte stillinger, som for eksempel professor II-stillinger.

Diskusjon

Målet med undersøkelsen er å belyse forskernes praksis med forskningsdata, få innspill til kompetansebehov og legge et kunnskapsgrunnlag for videre utvikling av retningslinjer for forskningsdata på BI. Resultatene i undersøkelsen har belyst problemstillingene angående hvilke forskningsdata finnes på Handelshøyskolen BI og hvordan kan Handelshøyskolen BI som institusjon kan bidra til forskernes arbeid med data. Resultatene gir innsikter i dagens praksis med hensyn til finansieringskilder, dataproduksjon, dataformater, lagringsmedier og syn på tilgjengeliggjøring av data. Respondentene gav innspill vil bli brukt til å avlede mulige tiltak til nytte for den enkelte forsker. Først vil resultatene diskuteres med hensyn til deres plass i en tematisk og forskningsmessig sammenheng, før det trekkes implikasjoner og anbefalinger om mulige tiltak.

Funnene støtter i hovedsak andre norske undersøkelser innen forskningsdata (L. H. Kvale, 2012; Norges forskningsråd, 2014a), men det kan se ut som BIs forskere vektlegger begrunnelsene litt annerledes på noen punkter. De øvrige undersøkelsene har hatt et større fokus på åpen tilgjengelighet av data. Denne undersøkelsen har med noen spørsmål som omhandler deling av data, men vektlegger i større grad muligheter for institusjonelle tiltak. Undersøkelsene er ikke direkte sammenlignbare på alle spørsmål, da kun et utvalg av spørsmål fra undersøkelsene er like, samt her er det i større grad åpnet for muligheten til å avgi mer enn ett svar på hvert spørsmål.

Undersøkelsen viser at 30 av respondentene (51 %) har i løpet av de siste årene brukt data fra andre forskere. Dette er en mindre andel enn 64 % som har brukt andres data i Forskningsrådets undersøkelse, hvor man bruker først data fra kollegaer ved egen institusjon, deretter fra internasjonale og nasjonale institusjoner. Fordelingen på BI er litt annerledes, hvor internasjonale institusjoner er mest brukt, nært etterfulgt av data fra kollegaer på BI og nasjonalt.

Hindre for deling av forskningsdata trekkes frem som en samling av faktorer, blant annet manglende intensiver for at en forsker skal dele data, å ha tid til å forberede data til arkivering, frykt for at andre skal misbruke data som legges ut, forskningsstrukturen er mangelfull.

Når det er snakk om hindre for deling av data, så fremhever flere av respondentene (34 %) at deling er vanskelig på grunn av at det kan redusere mulighetene for fremtidig publisering.

Dette burde indikere at skepsisen til tidsskriftenes krav om tilgjengelighet burde antas å være høy, også med tanke på muligheter for å gjenbruke data i forskjellige sammenhenger over tid. Det virker imidlertid som om forskerne er overveiende positiv til deling av data som en del av vitenskapelige tidsskrifter, hvor de fremhever mulighetene for å fremme av egen forskning (37 %), og kvalitetstegn for forskningen (25 %).

Forskningsrådets undersøkelse om syn på data som del av vitenskapelige publikasjoner viser en rangering fra først vekt på kvalitetssikring av publikasjoner, mer interessant for andre å følge ens forskning, positive konsekvenser for siteringshyppighet og ingen fordeler for den enkelte forsker. Undersøkelsen her legger først vekt på at publisering av data er med på å øke interessen for den enkelte forsker, kvalitetssikring av publikasjoner og ingen fordeler. 3 av respondentene på BI (5 %) har i liten grad tro på at publikasjon av data som del av vitenskapelige publikasjoner har en positiv innvirkning på siteringshyppighet. Dette skiller seg i stor grad fra Forskningsrådets undersøkelse, hvor en langt høyere andel 144 (11 %) mente at dette var positivt.

I likhet med Forskningsrådets undersøkelse (49,1 %) peker forskerne på BI at sikkerhet (38 %) er den viktigst faktoren til misnøye med eksisterende lagringsløsninger. Men på BI er kapasitet ved lagringsløsningene er like stor kilde (38 %) til misnøye, i motsetning til en langt lavere andel (21,8 %) hos Forskningsrådet. For mange lagringsløsninger er det litt mer misnøye med på BI (31 %), mot den nasjonale undersøkelsen (27,5 %). I undersøkelsen har svaralternativene fått inkludert alternativet for få lagringsløsninger, men det er få (6 %) som mener dette er et problem.

Forskerne bruker i størst grad de mest vanlige filformatene, slik som numeriske og tekstdokumenter. Dette er likt funn hos Forskningsrådet. I mindre grad anvendes webformater, billed-, video- eller lydfiler. Dette tyder på at enda er det ikke stor bruk av filformater som regnes å være på fremvekst i forskningsmiljøene, for eksempel mer bruk av bilder og webgenererte filformater. Dette kan sees på som interessant på grunn av OECDs mål om å bruke data på nye måter og med nye forskningsmetoder.

Forskerne benytter flere lagringsløsninger for data etter avsluttet prosjekt, gjerne på flere lagringsmedium. Undersøkelsen fanger ikke opp årsaken til praksisen. Årsaken kan være knyttet til tanken om enkel tilgjengelighet, hvor fleksibilitet med hensyn til å ha data tilgjengelig fra flere kilder kan spille inn. En annen forklaringsmodell kan være behovet for å

sikre data ved å lagre data i flere medium. Tendensen er klar med tanke på bruk av dataarkiver, hvor det i meget liten grad benyttes seg av dette tilbudet. Forskningsrådets undersøkelse viser at 11 % lagrer hos et dataarkiv, mens i denne undersøkelsen utgjør dette 7 %.

Dersom forskerne skal se bort fra restriksjoner, så legger 90 % av respondentene vekt på at de gjerne vil dele data med andre. Primært vil data deles med personer i tilknytning til forskningsgruppen, sekundært med forskerne i organisasjonen internt og tilslutt med eksterne forskere. Forskerne ønsker i liten grad å dele data med andre uten en forskningsmessig tilknytning, slik som sponsorer (19 %) eller allmennheten (20 %). Dette kan tyde på et behov for delingsløsninger internt og mellom forskningsinstitusjoner som de viktigste tiltak i første rekke.

Muligheter til profilering av egen forskning (37 %) og kvalitetssikring (25 %) trekker respondentene frem som et viktig element når det er snakk om hvilken betydning tidsskrifters krav om arkivering har. S. Kvale (1999) peker på at tilbakemeldinger fra kollegaer er en viktig faktor ved mesterlære. Evaluering blant fagfeller er et viktig element i forskersamfunnet, slik resultatene kan antyde. Et element av selvhevdelse og konkurranse kan også spores. Forskersamfunnets påvirkning og mulighetene for publisering i tidsskrifter synes å være det sterkeste insentivet for å ta vare på data i dag.

Norges forskningsråd er den største forskningsfinansiøren blant respondentene i undersøkelsen. Siden krav om DMP ble innført i 2014 vil få av respondentene ha erfaring med dette. Gjennom kommentarene peker 18 (59 %) av respondentene på at de ønsker støtte til å møte krav fra forskningsfinansiører om data og datadeling. Imidlertid, vil respondentenes finansiering ved svartidspunktet antyde at 20,6 % vil ha behov for datahåndteringsplaner om de skulle søkt midler nå.

I de videre vil de teoretiske perspektivene og overordnende forskningspolitiske bakgrunnen diskuteres i relasjon til resultatene. De forskningspolitiske føringer legger vekt på å kunne løse globale problemstillinger, samarbeid og nye måter å løse problemer på (Organisation for Economic Co-operation and Development, 2007) . Det vil trekkes inn tanker om mulige veier for videre forskning.

En av de viktigste forutsetningene for de forskningspolitiske føringene er at deling av data skal skape ny kunnskap gjennom nye måter å kombinere data på. Det er forventet at

analyseteknikker og behov for stor lagringskapasitet vil berøre disse fagområdene etter hvert som fagfeltene utvikler seg i retning av «big data» .

Resultatene angående filformater og datas innsamlingsmetode tyder på at i dag kombinerer forskerne ulike metoder, men i hovedsak er dette knyttet til tradisjonelle forskningsmetoder. Bruk av registerdata er relativt utbredt, men forskning som anvender «nye» filformater og bruk av beregningsvitenskap er mindre. Utvalget er noe skjevt med hensyn til at flere forskere med lengre erfaring har svart på undersøkelsen. Med tanke på om alder har sammenheng med deling av data, så vites ikke om erfaring er knyttet til bruk av flere kilder eller kombinasjoner av forskningsmetodiske fremgangsmåter.

Flere kilder fører til innovasjon er en annen antakelse knyttet til deling av deling av forskningsdata. Resultatene viser at metodetriangulering og bruk av flere datakilder anvendes, men man vet ikke om dette er en del av historiske/longitudinelle studier på samme data, eller forskningssubjekter, eller om dette fører til innovasjon. Det er heller ingen elementer som peker på i hvor stor grad forskningen er i front innen egne eller tverrfaglige områder, og om infrastrukturen eller deling av data er en medvirkende årsak til fremdragende forskning. På den måten kunne det vært interessant å se på praksis hos Senter for fremdragende forskning omkring deres datapraksis og tilgjengeliggjøring av data om dette skiller seg fra øvrige norske forskningsinstitusjoner.

Tematisk, så har undersøkelsen relativt stort fokus på deling av data og forskernes syn på dette. Undersøkelsen ser på hvordan forskere bruker forskningsdata i dag. Det fokuseres ikke på hva som forskerne eller forskningsgrupper antar forskningsmetodologisk utvikling på eget fagfelt vil være. På denne måte vil undersøkelsen i begrenset grad kunne bidra til innspill til strategiutvikling på lang sikt i organisasjonen. En mulighet er å supplere med kvalitative intervjuundersøkelser av fagpersoner angående endring av forskningsmetodologi innen eget fagfelt.

En annen antakelse som ligger til grunn for argumentasjonen for tilgjengeliggjøring av forskningsdata fra myndighetenes side er at synliggjøring av forskeren vil føre til samarbeid (Organisation for Economic Co-operation and Development, 2007). Undersøkelsen viser at forskerene deler data, og vil gjerne dele deres data med andre forskere. Men dette foregår utenfor de etablerte forskningsdataarkivene. Data publiseres først og fremst som følge av krav fra forlag ved publisering av vitenskapelige artikler. Dette kan henvises til en av de tre faktorene for informasjonsdeling (Pilerot, 2013), slik at forskeren handler etter materielle og

økonomiske forutsetninger ved publisering. I undersøkelsen uttaler forskere seg om hvem sine data som man bruker, men undersøkelsen fanger ikke opp hvordan man får informasjon om hvilke data som eksisterer. Den sier således ikke noe om hvordan kontakt mellom forskere oppstår, eller hvordan data utveksles. En av respondentene uttaler at «...andre kan ta kontakt på epost», men det er en forutsetning at man har kjennskap til hvilke data andre har. Pilerot (2013) har tidligere pekt på at personlige kontakter og tillit til den enkelte forsker er viktig for hvem som kontaktes for å dele, eller hvem som data man ønsker å benytte. Han mener dette er et brudd på forskningens manglende nøytralitet, disinteressethed, slik som (Merton, 1957) fremsetter det. Dette kan være en mulig forklaringsmodell for den manglende bruk av arkiver på det nåværende stadium og tilgang til data. Organisation for Economic Co-operation and Development (2007) legger til grunn at deling av data fører til samarbeid, men per dags dato er heller forklaringsmodellen mer andre veien, slik at samarbeid fører til deling. Ved en bedre tilgang på data via arkiver slik som finansierer og myndigheter legger opp til, vil dette i fremtiden føre til en endring. Slik at kontakter og kjennskap som kan være en forutsetning i dag vil endres over tid? Forskerne er interessert i å vite mer om mulige kilder for data, da å få hjelp til å finne tilgjengelige kilder er et av deres ønsker til tiltak.

Respondentene peker selv på at publisering av data i tidsskrifter er et synliggjøringskriterium i seg selv og en stadfesting av forskningens kvalitet. På sikt er det et spørsmål om data vil brukes som en synliggjøringsmekanisme for forskningsinstitusjoner. Ut fra tanken om at publisering av data er en kvalitetsfaktor på lik måte som vitenskapelige publikasjoner kan være dette.

Mertons normer for vitenskapelig etos (Gilje & Grimen, 1995) legger til grunn et felles eie av vitenskapelige resultater gjennom deling. Dersom man legger vekt på Mertons normer angående deling av resultater som måten å bringe forskningen fremover, så burde det vært enighet blant respondentene om deling av data. Den store viljen til å dele til forskere dersom restriksjoner ikke inntreffer tyder på dette. Respondentene fremhever forskningsetiske aspekter som det største hinderet for deling av data. Spesielt anonymisering fremheves, som igjen avspeiler normer i forskningsmiljøet og opplevde forpliktelser over forskningssubjektene.

Forskeren har rett til å beholde originalitet for eget arbeid gjennom første publisering, slik at man får anerkjennelse fra eget arbeid er en av Mertons normer. Dette står sterkt hos respondentene, slik som også ble funnet hos L. H. Kvale (2012). Samtidig, så viser

respondentenes argumenter for hvorfor data skal lagres over tid til Mitrofs motnormer om å beholde data mer eksklusivt til man har fått anledning til å publisere resultater fra dem.

Videre er det en henvisning Mertons normer om felleseie, etterprøvnbarhet og akademisk redelighet ved sitater som «allow other researcher to replicate and build on my research...». Samtidig som det finnes faglige kriterier med henhold til “brukstid” for data /varighet, slik at dette kan reflektere standarder innen eget fagfelt. Forskningens ytre rammefaktorer er inkludert i respondentenes kommentarer ved henvisning til at tidsskriftene etterspør “journals usually require this” “required by journals in my field».

Undersøkelsen peker på at forskerne opplever en frykt for at innhold kan misforstås. S. Kvale (1999) redegjørelse for mesterlære peker på at det kan være kulturelle eller sosiale forskjeller innen disipliner eller at det kan være flere praksiser innen et felt. Noen forskere ser lite eller ingen problemer med deling av data, mens andre ser dette som et problem. Kan det finnes konkurrerende normer innen et miljø, eller subkulturer, er det fag- eller metodisk tilnærming som definerer – eller begge deler? Veien videre med hensyn til forskning bør i høyere grad belyse dette. Forskningsmessig kan det være interessant å se om inndeling i enda mindre grupper med bestemte faglige fokus vil vektlegge andre synspunkter eller behov. For en utdanningsinstitusjon vil dette være vanskelig å få gjennomført i praksis, da målet er å fange opp alles behov og ta hensyn til bruk av alle forskningsmetoder.

Lagring av data innebærer kostnader. Kunnskapsdepartementet (2013) har i stor grad fokus på investeringer ved datalagring, men foreløpig er ikke kostnader noe som forskerne er i stor grad opptatt av. De er fornøyd med dagens løsninger, men bemerker samtidig at det mangler infrastruktur. Behovene som de peker på er først og fremst rettet mot lokal infrastruktur og deling innen egen institusjon. Det eksisterer en rekke nasjonale tiltak som forskningsinstitusjonen kan benytte seg av. Det er mulig at anvendelsen av tjenestene for lagring av data og bruk av dataarkiver kunne ha gunstig innvirkning ved å samspille med lokal infrastruktur.

Forskerne kan få økonomisk støtte gjennom myndighetene og forskningsfinansiørens tilrettelegging for å søke midler til tilrettelegging og deling av data. Forskerne synes mer opptatt av hjelp til å søke forskningsmidler på generell basis. Dette kan skyldes at dette er en relativt ny ordning, men det kan ha andre årsaker. Undersøkelsen viser at respondentene har en bevissthet om at data kan ha en begrenset verdi. Å ha et bevisst forhold til datas verdi og bevaringslengde er en av bakgrunnene for at finansiører krever en datahåndteringsplan.

Resultatene i denne undersøkelsen viser at bevaringsverdien som forskerne gir data er lavere enn finansierens krav om langtidslagring. Dette utløser et interessant tema, som kan være aktuelt for videre studier. Er det et skille mellom det som forskerne selv vurderer som bevaringsverdig og det som finansierers eller nasjonale dataarkiver mener bør lagres over tid?

I resultatene henvises det lite til økonomisk vinning som hindring for å dele data. Tid og eksklusiv rett til å publisere på egne data i forkant, slik som motnormene henviser til synes å stå sterkere.

Det er et spørsmål om lagring av data i seg selv vil føre til mer bruk av andres data. Dette kan bety at studier som kan følge opp forskernes nettverk og relasjoner kan være aktuelt som tema ved fremtidig forskning. Vil forskernes nettverk påvirkes ved å tilgjengeliggjøre data, og vil det i større grad føre til tverrfaglige kontakter slik som OECD legger opp til?

Nettverksanalyser og siteringsanalyser kan muligens brukes til å fange opp dette, dersom det er et kriterie.

Respondentene fremhever selv at bevaring av data er viktig med tanke på å kombinere data på nye måter og anvende dem i longitudinelle studier. Når data kombineres fra flere kilder kan eierskap og lisenser føre til restriksjoner for videre deling, slik som Kruse og Thestrup (2014) fremhever. Vil dette være et element som i stor grad skaper begrensninger ved krav om tilgjengeliggjøring av data gjenstår å se.

Med tanke på at i Norge er arbeidet med forskningsdata et nytt felt, vil det for videre forskning være mulig å se på endringer over tid på feltet. Organisation for Economic Co-operation and Development (2007) skisserer at det kan oppstå nye stillingskategorier og endrede arbeidsoppgaver, som nye typer forskning og spesialiseringer. Spesialisering med hensyn til støttefunksjoner hvor dette berører både teknisk utvikling og lisensiering vil interessant å følge. Foreløpig synes det å være et godt stykke igjen før de forskningspolitiske målene blir en realitet gjennom forskernes praksis.

Endringer i spørreskjema

Metodologien anvendt i undersøkelser utledet i Data Asset Framework (Jones et al., 2009) har gått grundig og detaljert til verks. Denne tilpasningen til norske forhold har redusert antall spørsmål i undersøkelsen betydelig, for eksempel i forhold til fremgangsmåten anvendt ved University of Southampton (McGowan & Gibbs, 2009). Erfaringene tyder på at ytterligere

endringer kan gjøres for å få frem forskernes behov og hvordan forskningsinstitusjonen kan bidra til dette.

Undersøkelsen har en lav svarprosent 18,15 %. Dette er ikke et uvanlig problem ved spørreundersøkelser (Fink, 2013). Et tiltak kan være bruk av «personlige invitasjoner». Dersom institusjonen åpner for det kan lenker i invitasjoner til spørreundersøkelsen være koblet til uttrekk av deskriptive opplysninger i personalsystemet, eks. kjønn, alder o.l. Dette ville gitt data et bredere analysegrunnlag. Korte undersøkelser er en fordel for å få en høyere svarprosent. Undersøkelsen inneholdt maksimalt 22 spørsmål. Det var estimert at undersøkelsen ville ta ca. 15 min å besvare, imidlertid viste det seg at snittiden var 7-8 minutter. Dette er en betydelig reduksjon i forventet tidsbruk som kan være utslagsgivende i forskerens tidspressede arbeidsdag. Bruk av det nettbaserte verktøyet Qualtrics (2015) var tilrettelagt for mobile plattformer, som må antas å være positivt for tilgjengeligheten av undersøkelsen.

Spørsmålene veksler mellom lukkede og åpne kategorier av svar, samt muligheter for å velge flere svaralternativer på spørsmål. Utstrakt bruk av å velge flere svaralternativer på et spørsmål tyder på forekomst av mange mulig konstellasjoner og varianter som er med på å gi et intrikat bilde av forskningsdata og sammenhenger ved videre studier.

Undersøkelsen består av et lite utvalg på et bestemt tidspunkt i en bestemt kontekst, slik at det er vanskelig å generalisere. Resultatene viser en avgrenset kontekst innen en organisasjon, slik som var intensjonen med undersøkelsen. Resultatene viser at respondentene har en overvekt av personer med lengre erfaring. Andre studiet har funnet at alder påvirker deling av data (Andreoli-Versbach & Mueller-Langer, 2014). I dette tilfellet kan det stilles spørsmål med om noen grupper definerer seg selv ut som respondenter på grunn av manglende eller liten erfaring med temaet. På denne måten kan det være skjevheter i utvalget. Det kan være en potensiell feilkilde, siden hensikten var å nå alle vitenskapelige ansatte i organisasjonen.

En annen faktor kan være at noen respondenter definerer seg ut av undersøkelsens målgruppe basert på forskningsmetodene som de anvender. Et aspekt kan være at man ikke produserer data selv, slik at deling og arkivering ikke er aktuelt. Det kan tenkes at forskerne som anvender kvalitative data har en mer skeptisk holdning til deling av data, slik at de anser de forskningsetiske aspektene som uaktuelle for å dele data, og dermed uaktuell for deltakelse i undersøkelsen.

En mulig forbedring er å spisse invitasjoner og undersøkelsen slik at man klargjør for mulige respondenter at de er målgruppe. Et annet alternativ kan være å vektlegge inndelingen i forskningsmetode sterkere, slik at undersøkelsen oppleves som mer skreddersydd for respondentene. I undersøkelsen ble National Science Board (2005) sin inndeling brukt, hvor respondentene valgte mellom variablene observasjonsmessige data, registerdata, dokumentkilder, eksperimentelle data og beregningsvitenskapelige data. En finere inndeling i variabler kan anvendes for å skille de ulike typene fra hverandre. Dette er imidlertid en finere inndeling enn når dagens 4 paradigmer for vitenskapelig forskning (Bell et al., 2009) omtales. Paradigmene kan anvendes for å dele undersøkelsen inn i 4 «stier» for metodologisk fremgangsmåte. Man må imidlertid være klar over at mange av forskningsprosjektene anvender mer enn en metodisk fremgangsmåte, men dette burde være mulig å ivareta.

Et ytterligere alternativ kan være å rendyrke undersøkelsen ved etterspørre forskere innenfor forskningsfinansierens målgruppe med egenproduserte, digitale data. Dette synes å være et dårligere alternativ, dersom en slik undersøkelse skal være en del av et institusjonelt arbeid. Fra et institusjonelt perspektiv bør man ivareta alle forskere ved en institusjon. Forskere som ikke produserer data selv kan oppleve utfordringer med hensyn til å få tilgang til kilder eller registerdata, slik som noen av respondentene beskriver.

Fra et institusjonelt ståsted kan det være interessant å vite flere detaljer om hvilke kilder forskerne anvender når de ikke produserer data selv, både med hensyn til opplæring og markedsføring omkring tilgang til kilder.

En annen vri som kan supplere undersøkelsen er å se på hvordan forskerne utfører datalagring i praksis. En undersøkelsesvariant som DAF (Jones et al., 2009) foreslår er å få tilgang til filstrukturer som forskerne bruker, slik at man kan få en mer detaljert innsikt i praksis om hvordan data lagres. Funnene i undersøkelsen indikerer at de fleste forskerne lager data på egen PC (58 %), på bærbare lagringsenheter (36 %), eller på institusjonell server (36 %). Bruk av internettbasert lagring som skytjenester (31 %) er utbredt. Hvordan respondentene reflekterer omkring egen PC kan i undersøkelsen er usikkert, dette kan være en PC som man har gjennom sitt jobbforhold eller egen eid PC. På BI får ansatte en bærbar PC hvor de kan lagre direkte på harddisken eller på en egen katalog på et fellesområde. Det er usikkert om respondentene gjør denne fininndelingen når de besvarer skjemaet, eller om de er klar over forskjellene for lagring. Lagres data på egen jobb PC på personlig fellesområde vil det for

eksempel tas back-up av alle data, men det vil ikke gjøres dersom det lagres direkte på PC'ens harddisk.

Undersøkelsen har noen forbedringspotensialer gjennom å aktivere målgruppene til å delta i undersøkelsen, fange opp respondenter som eventuelt definerer seg selv ut av undersøkelsen eller få frem andre aspekter som er interessant for et institusjonelt arbeid med filer og filstrukturer og datakilder. En annen forskningsmetodisk fremgangsmåte kan brukes for å utforske andre tematiske områder. Kvalitativ metode kan anvendes for å intervju forskningsgrupper, ev. fokusgruppeintervjuer om deres behov og praksis som kan berike eksisterende undersøkelse. Imidlertid kan bruk av kvalitativ metode for institusjonelle kartlegginger ansees å kreve mer tidsbruk for forskerne som informanter, men få frem en annen dybde på svar og utveksling av informasjon gjennom praksis.

Anbefalinger

Det er i kapittelet om forskernes syn på lagring av forskningsdata foreslått en rekke forskjellige tiltak for arbeid med forskningsdata, både på lokalt, nasjonalt og internasjonalt nivå. Tiltak på lokalt nivå bør sees i nær sammenheng med øvrige tiltak på nasjonalt, eventuelt internasjonalt nivå. Anbefalingene tar for seg ulike elementer og tiltak, som kan gjennomføres på kort eller lang sikt ved Handelshøyskolen BI. Noen av aktivitetene kan iverksette relativt enkelt og uten store kostnader, mens andre aktiviteter er av en mer langvarig karakter og kan kreve investeringer.

Flertallet av respondentene ønsker ulike former for tiltak på BI. Bedre infrastruktur, mer ressurser og opplæring som initiativ fra BI fremheves. Blant tiltakene som de ønsker initiert er bedre lagring og back-up under arbeid med data, hjelp til å finne og få tilgang på data og vite om «best practices». Respondentene som er misfornøyd med dagens lagringsløsninger peker på sikkerhet, kapasitet og manglende brukervennlighet.

I det videre foreslås mulige tiltak for infrastruktur, informasjonsaktiviteter, markedsføring av eksisterende tilbud, støttetjenester, ulike former for opplæring, diskusjoner om "beste praksis" innen forskningsmiljøene og opprettelse av koordineringsinstans/arbeidsgruppe.

Infrastruktur

Respondentene ønsker en bedre infrastruktur. Infrastruktur kan innebære lagringsløsninger og arkiver, eller tilgang på databaser eller annet materiale som kan anvendes som datagrunnlag for analyser.

Sikkerhet er et viktig element for forskerne, slik at data som legges ut ikke vil misforstås eller manipuleres. En rekke interesseorganisasjoner og tiltak har utledet hvilke karakteristiske arkiver bør ha for å hindre misforståelser og manipulering, gjennom funksjonaliteten i arkiver, sikkerhet og koblinger mellom kilder som vil bidra til å gi data en merverdi ved at informasjonen bidrar til forståelse for data.

Krav til arkiver hvor forskerene skal lagre og bevare data følger disse standardene, på denne måten vil kravene fra forskningsfinansiererne omkring funksjonalitet for dataarkiver oppfylles. Dette innebærer at en rekke arkiver anskaffer sertifiseringer om at de oppfyller funksjonaliteten. Data Seal of Approval er en anvendt standard (Dillo & De Leeuw, 2014). De nasjonale norske arkivene, NSD-arkivet og NorStore Research Data Archive, er bevisst på å oppfylle standarder fra interesseorganisasjoner og kriterier for Data Seal of Approval (Skagen, 2015).

Forskningsprosjektene genererer ikke store datamengder (under 500 GB). Dette kan tyde på at per dags dato er behovet for spesielløsninger for store datamengder lite. Dette impliserer at det trengs en IT-infrastruktur som kan ivareta behov for de fleste brukerne. Tilfredsheten med eksisterende lagringsløsninger antyder at de fleste forskerne kan benytte seg av eksisterende infrastruktur.

Kapasitet for lagring av data er en av faktorene som respondentene er misfornøyd med. Dette kan innebære et behov for tilpasning for de gruppene som har større datamengder eller anvender databaser for egen forskning. Mulig evaluering av egne tilbud for databaser og store datamengder kan være aktuelt. Store datamengder kan henvises til andre nasjonale tilbud for lagring og tilgjengelig tilbud for utveksling av data, inkludert sensitive data.

Respondentene er opptatt av fleksibilitet og tilgjengelighet av filer. Utveksling av data mellom forskere eller ved bruk av dataarkiver vil kreve overføring av datafiler. De fleste vil være i stand til å benytte seg av Uninetts «File Sender» tjeneste. Tjenesten kan benyttes av institusjoner i høyere utdanning for å oversende filer opptil en størrelse på 1 000 gigabytes.

Respondentene peker på at det finnes mange løsninger for lagring, som kan indikere at det er en utfordring å velge den beste løsningen for arbeidsflyt og bevaring av data. Et mulig tiltak kan være å synliggjøre det eksisterende tilbudet med vekt på hvilke tjenester som finnes, hva som er fordeler og ulemper med de forskjellige løsningene.

Respondentene er fornøyd med dagens løsninger for data, men data deles i liten grad gjennom dataarkiver. Krav til arkivering vil sannsynligvis øke i forskerens bevissthet etter hvert som kravene til tilgjengeliggjøring av data blir mer praktisert. Bruk av dataarkiver innebærer et arbeid med å tilrettelegge data for arkivering gjennom beskrivelse av data, og dokumentasjon slik at deres data ikke misforstås av andre eller kan etterprøves. Respondentene bruker i dag mest filformater anvendt til numeriske data og tekstdokumenter. Dette kan indikere behov for støtte gjennom å kunne dokumentere datapraksis ved anvendelse av filformatene i programvare for kontorstøtte og analyseverktøy i forskning.

Respondentene anvender mest data fra internasjonale forskningsinstitusjoner, men likevel er det en etterspørsel etter hvor kan man finne data til bruk for analyser. Dette kan tyde på et potensiale for at flere forskere ønsker å bruke andres data, som kan fremmes ved å formidle hvor data kan lokaliseres. Blant tiltakene kan være å etablere informasjonstjenester om tilgjengelige datakilder og danne seg en oversikt over dataproduksjon, enten på lokalt eller nasjonalt nivå. Siden data fra kollegaer og andre norske forskningsinstitusjoner er data som gjenbrukes mest, er dette et veldig aktuelt tiltak. Eksisterende instanser som registrerer data eller publikasjoner, som NSD eller Cristin, kan være potensielle tilretteleggere på nasjonalt nivå.

Som nevnt tidligere, vet vi lite om hvordan data utveksles i praksis ved internasjonalt samarbeid. Dersom data utveksles ved filoverføringer på e-post, så kan sikkerhet ved filoverføring være et aktuelt tema for opplæring. Kryptering og beskyttelse av filer ved overføring på e-post eller gjennom programvare for dataanalyse som forskerne benytter kan være et fokusområde.

Andre tiltak kan være å få en bedre arbeidsflyt og tilgjengelighet av data, for eksempel peker noen av respondentene på avtalereguleringer for tilgang på registerdata. Et alternativ er å prøve å få til samarbeid i forkant av etablering av en nasjonal infrastruktur for elektronisk utveksling av registerdata, RAIRD (2015). I dette tiltaket arbeider NSD og SSB med å gi bedre tilgang til data. Når infrastrukturen er tilgjengelig fra 2017 vil den gi tilgang til anonymiserte data fra Folkeregisteret, Nasjonal utdanningsdatabase (NUDB),

Selvangivelsesregisteret, Arbeidsmarkedsdata, Forløpsdatabasen FD-Trygd. Dette kan være aktuelt da noen av kommentarene i undersøkelsen nevner utfordringer tilgjengelighet, kostnader og behandlingstid for registerdata med spesiell vekt på data fra Statistisk sentralbyrå.

Grunnen til at en del forskere synes arkivering er utfordrende skyldes at systemene er vanskelig å bruke. Brukervennlighet og funksjonalitet bør være et sentralt krav ved anskaffelser av datakilder, og brukertesting av systemer kan være et aktuelt tiltak eller innspill fra BI.

Mer ressurser

Mer ressurser kan innebære en økonomisk satsing i form av investeringer eller arbeidskraft. Investeringer i infrastruktur for å skaffe et datagrunnlag er nevnt, og et tiltak for tilgang til registerdata RAIRD er under utarbeiding. Personale som kan utnytte eksisterende løsninger og formidle dette til fagmiljøene på BI vil være sentralt med hensyn til tiltakene som forskerne etterspør.

Å finne datakilder kan innebære en krevende øvelse i gjenfinning av informasjon. I en norsk kontekst finnes tilgjengelige data i en rekke forskjellige arkivinstanser. En utfordring er at der ikke finnes en felles samling av ressurser som forskerne kan trekke veksler på når de ønsker å gjenbruke andres data. Forskerne har gjerne kjennskap til ressurser innen eget fagfelt, men man kan tenke seg at dette er fremkommet over lenger erfaring og kontakter i miljøet. Spørreundersøkelser og intervju materiale kan finnes blant annet gjennom Norske undersøkelser via NSD. Andre fagfelt som geovitenskap kan finne data via Norstore. Registerdata er spredt over tilgang via en rekke forskningsinstitusjoner som SSB, Skatteetaten og helseinstitusjoner. Arkivportalen ivaretar en oversikt over instanser som tilbyr historiske dokumenter. NSD ivaretar også data fra samfunnsvitenskapelige internasjonale eller multinasjonale undersøkelser. Andre internasjonale kilder kan være Verdens helseorganisasjon eller Verdensbanken. Per dags dato finnes det ikke en felles inngangsport for søking eller gjenfinning av data. Norske arkiver er i begrenset grad registrert i internasjonale oversikter for datakataloger. Den varierte registreringspraksisen kan innebære at informasjon eller støttetjenester for å finne data kan være aktuelt.

Datatilbydere har ofte krav om dokumentasjon fra forskerene om at hun/institusjonen vil ivareta tekniske og forskningsetiske aspekter dersom hun skal få tildelt data.

Avtalereguleringer og dokumentasjon av IT-sikkerhet ved egen forskningsinstitusjon etterspørres i mange tilfeller. Dette er ikke den eneste utfordringen for å få tilgang til data. Tilgang til data kan innebære en kostnad for forskningsprosjektet. EU og forskningsrådet legger til grunn at delte data skal gis gratis tilgang til eller ved kostpris. Likevel er det slik at kostnader for uttrekk og klargjøring av data kan innebære store kostnader for et prosjekt. Dette vil i mange tilfeller implisere et behov for å innhente eksterne forskningsmidler. Tilgang til andres data kan innebære at krav om arkivering av data i etterkant av prosjektet kan påvirkes av de avtalemessige reguleringer som er satt på data som er produsert av andre forskere eller offentlige instanser som gis ved tilgang. Dette kan således hindre deling i etterkant av prosjektet. Dette leder over til tiltak i forbindelse med opphavsrett og lisenser.

Opphavsrett og lisenser

Respondentene har vist at de ofte kombinerer data som de samler inn selv med andres data. Bruk av data fra ulike kilder virker inn på anledningen som forskeren har til å bevare data over tid, eller kunne dele dem med andre forskere. For å kunne oppfylle krav fra finansierer er det viktig å skille klart mellom hva som er primærdata og sekundærdata, samt å definere hvilke data som «må» arkiveres. Lisensiering/klargjøring av avtaler blir viktig siden respondentene i stor grad bruker egne data og kombinerer med data fra andre forskningsinstanser/kollegaer. Sletting av data kan være en forpliktelse ovenfor forskningssubjekter, og som er meldt inn til personvernombudet for forskning. Dette aktualiserer behovet for avtaler for å kunne benytte data i flest mulig sammenhenger og kunne benytte data over tid.

Klare ansvarsforhold med hensyn til hvem som har råderett for data er spesielt aktuelt ved prosjekter hvor det er samarbeid mellom flere forskere eller forskningsgrupper, eller når studenter utfører datainnsamling og bearbeiding. Fokus på temaet opphavsrett og lisenser vil være med på å belyse spørsmål som: Hvem eier data? Hvordan kan man sette begrensninger på bruk? Hvilke implikasjoner har det? I hvor stor grad kan man be forskningssubjekter om å få bevare data, som gjerne er sensitive, over lengre tid?

Resultatene viser at forskerene ønsker å dele, men først og fremst med andre forskere. Forskerene kan selv regulere data gjennom å anvende lisenser ved deling av data. Lisenser og

retningslinjer for data er ikke mye pekt på blant forskerne i undersøkelsen, slik at mer informasjon om standardiserte lisenser og eksempler på bruk kan være aktuelt.

Forskerene er mer opptatt av tidsskrifters/forlags policyer for krav om lagring av data og hvordan dette praktiseres. For forskerne kan innsikt i hvordan man kan få kunnskap om krav til datahåndteringsplaner gjennom nasjonale og internasjonale forskningsfinansiører, eller vurdering av publisering i datatidsskrifter eller tidsskrifter med krav om datalagring.

Informasjon som gir forskerne enkel informasjon omkring: Hvilke krav finnes hos forlagene? Hvordan kan man oppfylle disse? Hvilken finansiører krever dette? Sherpa Juliet⁹ er en database med oversikt over open access tidsskrifter og forskningsfinansiørers politikk, driftet av høyere utdanningsinstitusjoner i Storbritannia, som kan anvendes til formålet.

Mange av de nevnte forslagene har elementer om støttetjenester i form av informasjonstjenester eller opplæring. I avsnittet under vil det nevnes ytterligere noen forslag.

Opplæring

Faglige diskusjoner under paraplybegrepet beste praksis kan være nyttig for å formidle lure triks og forventninger hos kollegaer.

I undersøkelsen pekes det på deling av data som et kvalitetskriterium for egen forskning, diskusjoner i fagmiljøene kan være knyttet til faktorer for hvordan kvalitet defineres innen eget fagfelt med hensyn til deling av data, og dokumentasjon av data slik at disse ikke misforstås eller feiltolkes.

Et annet tiltak kan være knyttet til den praktiske utformingen av et forskningsprosjekt med hensyn til dokumentasjon av forskningsmetode i analysefasen av forskningsprosjektet. Et eksempel kan være å vise hvordan verktøy for datainnsamling og analyse kan brukes for å effektivisere dokumentasjon som kreves ved arkivering. Til dette kan man utnytte eksisterende verktøy forskerne i mange tilfeller kjenner, som for eks. NVIVO og SPSS. «How to guides» utviklet av DDC kan gi inspirasjon (Digital Curation Center, 2015).

Etter hvert som fokus på forskningsdata og finansiørerer krav om lagring av data har vunnet om seg, så er litteraturen om forskningsmetode i ferd med å inkludere kapitler om datahåndtering/administrasjon av forskningsdata. Dette kan tyde på en endring av

⁹ Sherpa Juliet: <http://www.sherpa.ac.uk/juliet/>

vektlegging av temaet, slik at dette kan være aktuelt som et element inn i undervisning om forskningsmetode.

En annen mulighet er å vurdere formalisering av opplæring, gjennom studiepoenggivende kurs i ph.d-utdanningen.

Skriving av planer for datahåndtering, DMP, er integrert i forskningsrådets retningslinjer og vurderingskriterier for finansiering. Forskningsrådet kan inviteres til å holde seminarer spesifikt på datahåndteringsplaner, eller som en del av skriving av finansieringssøknader generelt. For øvrig vil datahåndtering kunne inngå som en del av assistansen til de prosjekter som har behov for DMP som en del av finansieringssøknadene.

En annen mulighet er å invitere arkivinstansene som forskerne har anledning til å benytte, som for eksempel NSD-arkivet og Norstore til å informere om sine tjenester. Vekt kan legges på informasjon om muligheter for korttidslagring/ langtidslagring, hvilke tilganger forskerne kan sette på sine data, eller hvilke konsekvenser arkivering eventuelt har for forskerens synlighet i fagmiljøet.

Generell informasjon og praktiske element kan gis gjennom informasjon om hvilke hjelp å støtte man kan få og hvilke tjenester som eksisterer på BI allerede. Dette kan muligens fungere som en markedsføring og klargjøring ovenfor forskerne.

Arbeidsgruppe

Mange mulige tiltak og aktiviteter krever koordinering og et behov for å klargjøre ansvar i organisasjonen. Det kan være et behov for et samarbeid mellom de mange representanter som forventes å ha en aktiv rolle i administrasjon av forskningsdata på BI, slik som Avdeling for forskning og fagressurser, Forskningsadministrasjon, IT-avdeling og Bibliotek, eventuelt instituttledere/kontorsjefer ved avdelingene.

Respondentene i undersøkelsen pekte i liten grad på behov for retningslinjer eller standarder innen organisasjonen. Forskningsrådet/Kunnskapsdepartementet oppfordrer til at institusjonen skal lage sine egne retningslinjer. Hvordan en retningslinje kan utformes, finnes det eksempler på i den internasjonale universitets- og høyskolesektoren. Det er opp til institusjonen om de vil legge inn krav om å lage datahåndteringsplaner for alle prosjekter, eller pålegge rapportering eller arkivering av data fra prosjekter utført i organisasjonen.

Institusjoner kan innføre intensiver ovenfor forskerne, slik at data kan være et element som etterspørres i forbindelse med ansettelsesprosesser eller at det gis en markering/belønning ved arkivering av data. Arbeidsgruppen bør være aktivt for å ta stilling til en retningslinje for forskningsdata ved Handelshøyskolen BI.

Funn, implikasjoner og tiltak fra undersøkelsen

Avslutningsvis er det i tabell 19 trukket frem hovedfunnene i undersøkelsen. Hovedfunnene legger grunnlaget for forslag til anbefalinger ved videre arbeid med forskningsdata på Handelshøyskolen BI.

Tabell 19. Hovedfunn med forslag til anbefalinger og tiltak

FUNN	UTFORDRING /IMPLIKASJONER	MULIG TILTAK/AKTIVITET
Forskere på BI samler i stor grad inn egne data som basis for analyser, men kombinasjon av ulike datakilder er vanlig.	For å kunne oppfylle krav fra finansiører er det viktig å skille klart mellom hva som er primærdata og sekundærdata, samt å definere hvilke data som «må» arkiveres.	Lisensiering/klargjøring av avtaler blir viktig siden respondentene i stor grad bruker egne data og kombinerer med data fra andre forskningsinstanser/ kollegaer.
Filformater anvendt til numeriske data og tekstdokumenter er mest brukt blant forskerne.	Filformatene som anvender programvare for kontorstøtte og vanlige analyseverktøy i forskning genererer filformatene.	Dokumentasjon av datapraksis bør utføres ved bruk av vanlige filformater og analyseverktøy.
Forskningsprosjektene genererer ikke store datamengder (under 500 GB).	Per dags dato er behovet for spesialløsninger for store datamengder lite. Tilpasning for de gruppene som har spesialbehov kan være aktuelt.	IT-infrastruktur trengs som kan ivareta behov for de fleste brukerne. Store datamengder kan henvise til andre aktuelle lagringstjenester og tilgjengelig tilbud for utveksling av data, inkludert sensitive data. Eventuelle vurderinger av egne tilbud for databaser og store datamengder kan være aktuelt.
Forskere anvender mest data fra internasjonale forskningsinstitusjoner.	Potensiale for å bruke andres data er fremdeles høyt.	Informasjonstjenester om tilgjengelige datakilder. Lokal eller nasjonal oversikt over dataproduksjon.
Etter avsluttet prosjekt lagrer alle forskerne data. De fleste forskerne lager data på egne lagringsenheter, på egen PC eller på institusjonell server. Dataarkiver er i liten grad benyttet.	De forskjellige lagringsmåtene har sine fordeler og ulemper. Manglende lagring i dataarkiver gir liten mulighet for å fremme institusjonens forskning og oppfyller ikke finansiørens målsetninger.	Informasjon/opplæring om alternative løsninger for datalagring, fordeler med å lagre på institusjonell server, lagringsløsninger fra NSD og Norstore for NFR finansierte prosjekter. Praktisk hjelp til å lagre data i arkiv kan være aktuelt, sammen med diskusjoner om beste praksis i fagmiljøene. Klargjøringen av hva arkivering innebærer for den enkelte forsker/forskergruppe.
De fleste forskere er fornøyd med dagens lagringsløsninger.		
Forskere som er misfornøyd med lagringsløsningene viser til sikkerhet, kapasitet og manglende brukervennlighet.	Dette kan tyde på et behov for å identifisere forbedringspotensiale i eksisterende løsninger, eller fokus på alternative lagringsløsninger som er tilgjengelig.	Informasjon om hvordan data kan sikres, utforskning av muligheter til bruk av nasjonale tjenester for sensitive data. Beskrive for brukerne eksisterende kapasitet som er tilgjengelig på BI, og eventuelt hvilke alternative løsninger som kan benyttes. Klargjøring av hvilke utfordringer den enkelte forsker møter som er misfornøyd.

Tabell 20. Hovedfunn med forslag til anbefalinger og tiltak, forts.

FUNN	UTFORDRING /IMPLIKASJONER	MULIG TILTAK/AKTIVITET
Forskerne tar vare på data for å kunne gjenbruke data i nye analyser, dokumentere og etterprøve resultater.	Etterprøving og gjenbruk kan implisere et behov for å ha dokumenterte prosesser og analyser. Gjenbruk ved nye analyser impliserer vanligvis et tidsaspekt, gjerne over flere år. utfordringer med endringer/nye versjoner i programvare og utstyr som er brukt. Ivaretagelse av datas overføringsmuligheter til nye systemer og programvare kan være nødvendig.	Opplæring i dokumentasjon av data Formidling av fordeler med bruk av arkiv ved langtidslagring, jf. versjonshåndtering og oppdatering av programvare.
Anonymiserte data kan gjerne deles med andre i samme forskningsprosjekt og andre forskere utenfor BI, men i liten grad med allmennheten.	Anonymiseringer er den største restriksjon for deling av data.	Praksisdiskusjoner om når data er anonymisert godt nok. Opplæringstiltak om hvordan avtaler med forskningssubjekter/samtykke fra respondenter kan formuleres for gjenbruk av data over tid, gjerne i samarbeid med Personvernombudet for forskning / Regional etisk komité.
De største hindre for å dele data med andre forekomst av sensitive data, tid for å gjøre data tilgjengelig og manglende muligheter for fremtidige publikasjoner.	Hindre for deling av data er sammensatt av mange forskjellige aspekter, som karakterer ved data, forskerens arbeidssituasjon og faglige posisjonering/fagformidling. Slik er det er vanskelig å peke på en eller et par faktorer som vil være av utslagsgivende for deling av data	Diskusjoner i fagmiljøene om beste praksis, eks. når er et tema «utpublisert» og hva gjør «erfarne datapublisere», «hva er sensitive data, og på hvilket nivå er det akseptabelt å dele dem?»
Korttidslagring av data vurderes som mer aktuelt enn langtidslagring av data.	Vurdering av datas bevaringsverdi er et sentralt aspekt ved søknader om forskningsmidler, hvor det skal lages en datahåndteringsplan. Forskningsrådets kontrakter setter som krav langtidslagring, dvs. 10 år.	Opplærings- og praksisdiskusjoner om behovet for å lagre data over kortere eller lengre tid? Informasjon om tilbud om bevaring av data i et kortere tidsintervall enn 10 år. Vurdering av datas bevaringsverdi ved finansieringssøknader, hvor datahåndteringsplaner må fylles ut.
Bedre infrastruktur, mer ressurser og opplæring er ønskelig at BI iverksetter som tiltak.	Tiltakene som ønskes på BI er knyttet til infrastruktur og forskjellige varianter av støttetjenester, i forkant og under forskningsprosjekter. Bevaring av forskningsdata gjennom arkivløsninger er mindre prioritert.	Informasjon og støttetjenester, investeringer i infrastruktur for pågående prosjekter, samt opplæringstiltak og initiering av faglige diskusjoner for å belyse praksis.
Under forskningsprosessen ønsker forskerne bedre lagring og back-up under arbeid med data, hjelp til å finne og få tilgang på data og vite om «best practices».	Lagring av data under forskningsprosessen, tilgang til data og praksisutøvelse er viktig for forskerne.	Informasjon om IT-tjenester og opplæring i bruk av verktøy for lagring og deling av data. Støttesider /nettsider om datatilgjengelighet. Opplæring og diskusjoner om beste praksis.
Deling av data integrert i vitenskapelige publikasjoner er viktig for forskerens egen status i forskningsmiljøet.	Publikasjoner har stor betydning for forskerens faglige status, og er et kvalitetsmessig kriterium for forskeren. Vekt på hvilke publikasjoner som krever data, og hvordan man kan tilrettelegge for publikasjon av data kan være i fokus.	Forskingsformidling med vekt på hvordan man kan fremme din faglige status gjennom datapublisering. Finne eksempler fra fagmiljø om erfaringer og suksess ved deling av data.

Konklusjon

Arbeidet med undersøkelsen er en del av å skape et kunnskapsgrunnlag for strategiske og praktiske beslutninger om utvikling av forskningsdata innen en høyere utdanningsinstitusjon.

Resultatene i undersøkelsen viser forskernes innstilling til deling av forskningsdata, deres datapraksis og ønsker om forbedringer i arbeid med forskningsdata på en forskningsinstitusjon. Undersøkelsen bekrefter funn fra tidligere norske studier om deling av forskningsdata (L. H. Kvale, 2012; Norges forskningsråd, 2014a).

Forskerne er positivt innstilt til deling av data. De største hindre for å dele data med andre forekomst av sensitive data, tid for å gjøre data tilgjengelig og manglende muligheter for fremtidige publikasjoner. Anonymiserte data kan gjerne deles med andre i samme forskningsprosjekt og andre forskere utenfor BI, men i liten grad med allmennheten.

Forskernes datapraksis viser at etter avsluttet prosjekt lagrer alle forskerne data. De fleste forskerne lager data på egne lagringsenheter, på egen PC eller på institusjonell server. Dataarkiver er i liten grad benyttet. De fleste forskere er fornøyd med dagens lagringsløsninger. Forskere som er misfornøyd med lagringsløsningene viser til sikkerhet, kapasitet og for mange lagringsalternativer.

Forskere på BI samler i stor grad inn egne data som basis for analyser, men kombinasjon av ulike datakilder er vanlig. Forskere som anvender andres data anskaffer dem fra internasjonale forskningsinstitusjoner. Forskningsprosjektene genererer ikke store datamengder (under 500 GB). Filformater for numeriske data og tekstdokumenter er mest brukt blant forskerne. Det er oppmerksomhet om datas bevaringsverdi, hvor korttidslagring av data vurderes som mer aktuelt enn langtidslagring av data.

Forskerne tar vare på data for å kunne gjenbruke data i nye analyser, dokumentere og etterprøve resultater. Deling av data integrert i vitenskapelige publikasjoner er viktig for forskerens egen status i forskningsmiljøet.

Bedre infrastruktur, mer ressurser og opplæring er ønskelig at BI iverksetter som tiltak. Under forskningsprosessen ønsker forskerne bedre lagring og back-up under arbeid med data. Støttefunksjoner som hjelp til å finne og få tilgang på data er ønsket. Innsyn i hva som er den beste praksis på eget fagfelt er av interesse.

Utviklingen på forskningsdatafeltet kan på mange måter sies å gå fra et individualisert ansvar for den enkelte forsker/forskningsgruppe basert på forskningsetiske normer og verdier, til et mer mangefasettert område hvor finansierer, universiteter og høyskoler, samt myndigheter ønsker tilbakemelding og tilgang til data som produseres. Strategisk utvikling innen en forskningsinstitusjon påvirkes av myndigheter og forskningsfinansierer krav, hvorpå det dannes overordnede strategiske mål. Gjennom handlingsplaner settes tiltak for at man skal oppnå nye mål, og eventuelt endre praksis i organisasjonen. På denne måten kan organisasjonen bidra til å oppfylle myndigheter og finansierer pålegg om endring i forskernes praksis. For institusjon vil dette være et skjæringspunkt mellom myndighetenes krav og den enkelte forskers interesser, samtidig som institusjonen kan ha egeninteresse for profilering av sin egen posisjon gjennom forskernes arbeider.

Forskningsdata er et tema som berører alle forskere. Forskerne er opptatt av hvordan de kan ivareta data i løpet av forskningsprosessen. Det er imidlertid ikke alle forskere som søker om forskningsmidler, basert på praksis i Norge i dag er det kun en andel som er forpliktet til å lage en datahåndteringsplan. Å lagre data er forpliktende gjennom kontrakter som forskerne har med forskningsfinansiørene. Forskerne ser på metodologiske fremgangsmåter og forpliktelser ovenfor respondenter som viktige pekepinne for hvilke data som deles med allmennheten. I praksis vil kun en liten del av data deles, og være åpent tilgjengelig gjennom dataarkiver. Andelen forskere som deler data vil være høyere enn det som vises gjennom dataarkiver, da personlige kontakter synes å være en forutsetning for deling.

Forskningen på utøvelse av forskningsdata inneholder en rekke arbeidsprosesser for lagring og eventuelt deling av data. Undersøkelsen viste at feltet har mange forskningsmessige implikasjoner som gjør det vanskelig å ivareta behov gjennom få tiltak. Tiltak knyttet til diskusjoner av forskningsetikk og samhandling mellom forskere gjennom avtaler og reguleringer, praktiske og brukervennlige tekniske løsninger må til dersom utfordringer i prosessen med administrasjon av forskningsdata skal bidra til overordnede mål om deling av data.

Variasjonene i praksis og mangfoldet av muligheter betyr at en rekke kompetanser involveres slik at dette kan innebære et behov for en tydeligere avklaring av roller i sektoren. Det er en stor andel aktører som er involvert ved en forskningsinstitusjon slik som ledelsen og forskningsadministrasjon, forskere og forskningsgrupper, IT-avdeling og eventuelt bibliotek. Utenfor forskningsinstitusjon involverer arbeidet forskningsfinansierer, myndigheter og

arkiver for datalagring. Forskningsdatas nyvunne fokus innen forskningsinstitusjonene har ført til at hvordan dette arbeidet skal organiseres og koordineres mellom aktørene i sektoren er ikke avklart per dags dato. For forskningsinstitusjonen handler dette å ta beslutninger om forskningsdata på et best mulig kunnskapsgrunnlag for å redusere usikkerheten omkring beslutningene som tas på en rasjonell måte.

På lang sikt handler det om å påvirke praksis og holdninger hos arbeidstakere gjennom å sette fokus på forskningsdata på agendaen i faglige forum og supplere med tiltak for opplæring. Vitenskapelige ansatte i en forskningsinstitusjon er kunnskapsmedarbeidere formet av tradisjoner, faglige standarder og praksis innen eget fagområde. Forskere er vant til en selvstendig posisjon med liten grad av styring.

I forskningsinstitusjonen kan retningslinjer og krav om tilgjengeliggjøring av forskningsdata oppfattes forskjellig, basert på fagmiljøenes definisjon av data og eksisterende normer og verdier. Faglig kvalitet og anerkjennelse står sterkt i forskerens bevissthet.

Datapubliseringsaktiviteter satt inn en faglig kontekst vil være et sterkt insentiv. Kvalitet i forskningen er et fellestrekk av interesse for myndigheter, finansierer og forskere.

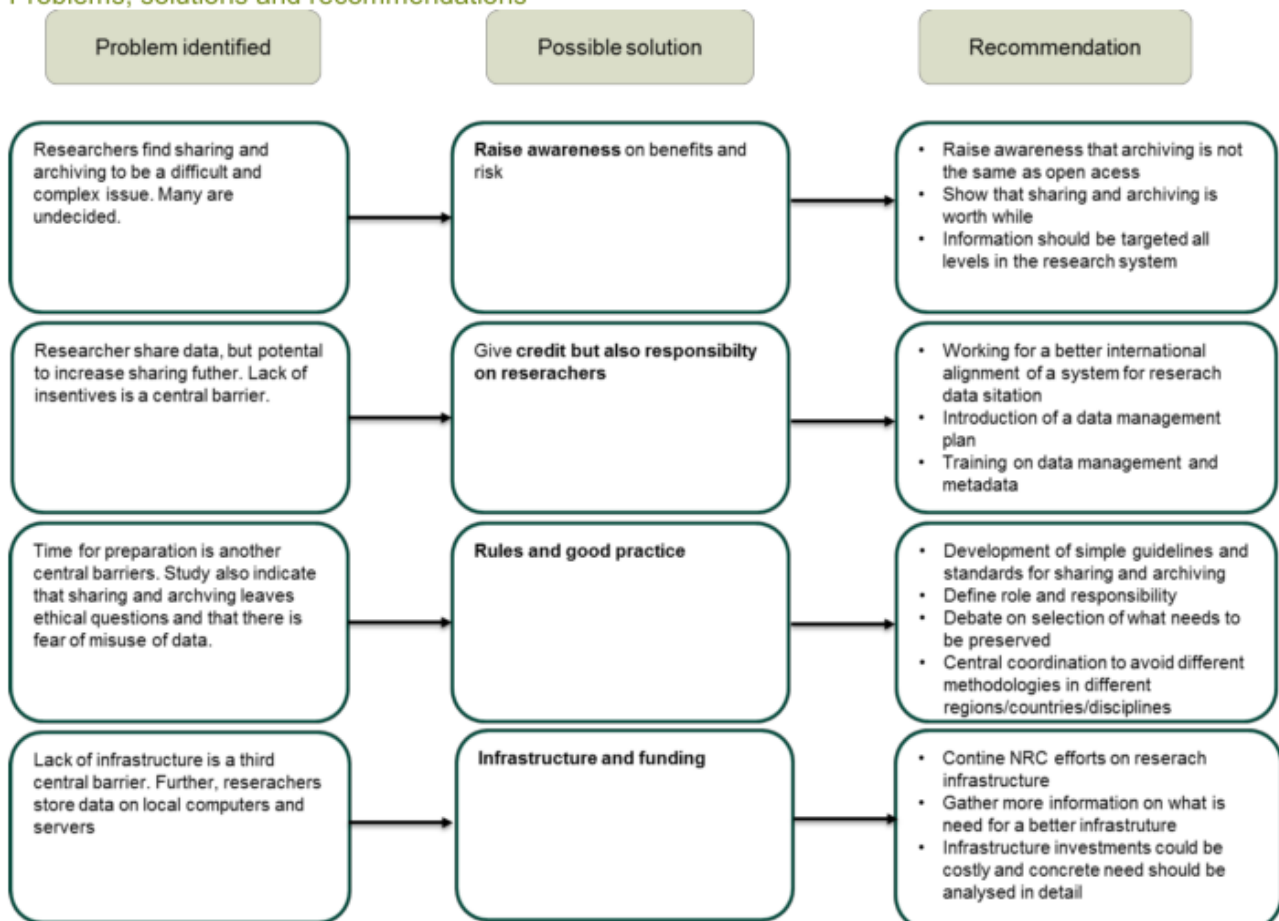
Studier av forskningsdata har hatt et oppsving etter myndigheters og finansierers fokus på tema. Fremdeles er det mange aspekter eller temaområder som kan ha nytte av videre forskning. Praksisteori kan brukes som utgangspunkt som et teoretisk rammeverk, men andre aspekter for videre forskning kan trekkes inn, slik som endringsledelse, strategiutvikling og kunnskapsledelse. Forskerens praksis med bruk av andres data. Evaluering av funksjonalitet i IT-tilbud og arkiver for å se hvordan de ivaretar forskningsdatatjenester til fagmiljøene. Etter innføring av opplæringstiltak og informasjon til forskerne som dataprodusenter og databrukere kan evaluering være aktuelt.

På lenge sikt gjenstår det å se om idealene som fremsettes av myndigheter og OECD om at deling av data er med på å gjøre forskningen bedre inntreffer. Vil forskningen ha en bedre kvalitet? Vil forskningen effektiviseres? Vil forskningen globaliseres og synliggjøres bedre? Vil forskerne samarbeide mer og skape ny kunnskap? Vil forskningsetisk praksis og behov for kontroll og sikre data, anonymisering og andre begrensninger føre til at deling av data sjelden vil forekomme?

Appendiks

Appendiks A. Foreslåtte tiltak fra Norges forskningsråd (2014 a).

FIGURE 1
Problems, solutions and recommendations



Source: DAMVAD

Figur 4. Problems, solutions and recommendations «Sharing Research Data» (Norges forskningsråd, 2014a)

Appendiks B. Invitasjonsbrev

You, as a researcher at BI, are invited to share your knowledge and opinions on research data. The aim of this survey is to identify advantages and disadvantages you face when working with research data. The purpose is to improve work with research data at BI.

This survey enquires about your research data practice:

- Acquire information about the research data you hold for the time being
- Your thoughts on data curation
- Your thoughts on how BI should enhance work processes for research data

Take the survey

(lenke til undersøkelsen)

You are able to answer open-ended questions in Norwegian. Completing the survey is estimated to max. 15 min.

All personal information collected will be treated confidential, and results anonymized when disseminated. This study will be a master thesis at Institute for Archive, Library and Information Science at Oslo and Akershus University College, and is scheduled for completion by December 2015. Project data will be deleted after project closure.

It is voluntary to participate in the survey, and you can at any time choose to withdraw your consent without stating any reason. The study has been notified to the Data Protection Official for Research, Norwegian Social Science Data Services.

If you have any questions concerning this project, please contact me:

Name

Contact information

Thank you for participating!

Appendiks C. Utsendt spørresundersøkelse

Research Data at BI Norwegian Business School

This survey enquires about your research data practice:

- Acquire information about the research data you hold for the time being
- Your thoughts on data curation
- Your thoughts on how BI should enhance work processes for research data

You are welcome to write open ended questions in Norwegian.

BEFORE STARTING A PROJECT

Think of one of you current or finished projects

Grant 1. Is this project grant funded?

- Yes (1)
- No (2)

Svar If Is this project grant funded? Yes Is Selected
Funding 1a. If yes, who is the funding agency/source?

- Norwegian Research Council (1)
- Governmental/ public agencies (2)
- Industry/ business (3)
- Private foundation / organisation (4)
- BI Norwegian Business School (5)
- EU programs (6)
- Other Norwegian Source (7) _____
- Other International Source (8) _____

2. What origin has data used as basis for your research analysis? (Check all that apply)

- My data are observational data, i.e. interviews, survey, instrument data (6)
- My data are register data, i.e longitudinal public, health data (10)
- My data are computational data, i.e. computer modelling, simulations (7)
- My data are experimental data, i.e resulting from experimental prosedures (8)
- My data are records, i.e. documents, movies (9)
- Other (please specify) (5) _____

DURING A PROJECT

3. Choose all of the following formats that best describe your research data (examples of specific file extensions are included):

- Numerical scores (.xls, .csv, .sav, .R) (1)
- Textual records (.doc, .rtf, .txt) (2)
- Images, sounds, videos and graphics (.jpg, .mov, .avi, .mp4) (3)
- Geographical information files (.gpx, .kml, .kmz) (4)
- Networked based data and web files (.html, .xhtml) (5)
- Databases (.db, .mdb, .sql) (6)
- Scripts and programming (Java,C++) (7)
- Other (please specify) (8) _____

4. Indicate the approximate amount of data the project is expected to generate

- 1 - 500 gigabytes (GB) (1)
- > 500 - 1000 GB (2)
- 1 - 500 terabytes (TB) (3)
- > 500 - 1000 TB (4)
- 1 - 500 petabytes (PB) (5)
- > 500 PB (6)
- Don't know (7)
- Other (please specify) (8) _____

5. Have you within the last three years used research data gathered by other researchers?

- Yes (1)
- No (2)

Svar om 5. Have you within the last three years used research data gathered by other researchers?
Yes Is Selected

5a. Whose research data have you used the most within the past 3 years? (Multiple answers allowed)

- Research data from other researchers at my institution (1)
- Research data from other researchers at national institutions (2)
- Research data from other researchers at international institutions. (3)

AFTER A PROJECT

6. What is the most common way of archiving your research data after results are ready or beyond the life of a project?

- Portable storage unit (i.e.,USB flash drives, CD/DVD, External hard drive) (1)
- Hard drive of the instrument which generates the data (2)
- PC hard drive (3)
- Institutional server (4)
- Internet-based storage (e.g., cloud or grid storage) (5)
- Data is submitted to digital archive centre in my organisation (6)
- Data is submitted to a national digital archive centre (7)
- Data is submitted to an international digital archive centre (8)
- Do not archive (10)
- Don't know (11)
- Other (please specify) (9) _____

Keep 7. What applies for most of your research data – Identify for how long you plan on keeping data?

- Less than 3 years (1)
- 3 - 6 years (2)
- 7 - 10 years (3)
- 11 - 20 years (4)
- More than 20 years (5)

8. Please explain why you plan on keeping the data for this amount of time:

9. If your research data gets lost, how easily can you recreate it?

- Not possible (1)
- Hardly (2)
- With same effort (3)
- Very easily (4)
- Don't know (5)

10. Identify all of the following who you would like to share anonymized data with, if you had the ability to do so (check all that apply):

- Researchers, students, and staff working on the project (1)
- Other members of the affiliated School or Research Center (2)
- Researchers at other institutions (3)
- Project sponsors (4)
- General public (5)

11. Are you satisfied with your archiving solution?

- Yes (1)
- No (2)

Answer If 11. Are you satisfied with your archiving solution? No Is Selected

11a. If “not”, why are you not satisfied with your archiving solution? (Multiple answers allowed)

- Too complicated to use (1)
- Too expensive (2)
- Too little capacity (3)
- Too many archiving solutions (4)
- Too few archiving solutions (5)
- Not secure enough (7)
- Other (please specify) (6) _____

12. Do you see any challenges in making more of your research data available for other researchers?

Maximum 3 answers

- Preparing data for open access takes away valuable time for research (1)
- Lack of technical infrastructure (2)
- Reduce possibilities of future scientific publications (3)
- I am afraid other researchers will not understand my data (4)
- I cannot give access due to sensitivity issues (5)
- I cannot give access due to shared ownership (6)
- I am afraid data will be misused (7)
- I cannot give access due to intellectual property rights (8)
- It would be unethical (10)
- Open Access to research data might have a negative economic impact for me and my institution (9)
- I cannot give access due to commercial issues (11)
- I do not believe my research data is of interest to others (12)
- I do not believe data is secure at a data centre, journal site or alike (13)
- I do not know (14)
- Other (please specify) (15) _____

13. Do you welcome the trend of making data available as a part of scientific publications?

- Yes, it could mean that my research could be more interesting for others to follow (1)
- Yes, it is a sign that my research can be quality assured (2)
- Yes, it could mean that my data and or my publications will be more cited (3)
- No, I see no benefit for me (4)
- I do not know (5)
- Other (please specify) (6) _____

14. What efforts do you think BI should initiate to improve preservation, curating and sharing research data? (maximum 3 answers)

- Better infrastructure for research data (1)
- Implementation of a system for citation (2)
- More resources allocated to research data activities (3)
- Implementation of guidelines (4)
- More training on the topic research data (5)
- Implementation of standards (6)
- Make open access to research data an indicator in the funding scheme (7)
- I do not know (8)
- Other (please specify) (9) _____

15. Identify which of the following services might be useful in regards to the management of research data (check all that apply):

- Assistance to meet data sharing or data management requirements of funding agencies (1)
- Information about developing a data management plan (2)
- Information regarding data management best practices (3)
- Assistance finding and accessing data resources (9)
- Assistance using tools for information gathering and analysis (13)
- Data storage and backup while research is on-going (6)
- Tools for sharing research data as research is on-going (5)
- Assistance selecting data to preserve for the long-term (4)
- Assistance applying metadata to research data (8)
- Provide long-term data access and preservation (7)
- Information about citing data resources (10)
- None of the above (11)
- Other (please specify) (12) _____

16. Identify the department at BI to which you belong

- Accounting, auditing and law (1)
- Communication and culture (2)
- Economics (3)
- Finance (4)
- Innovation and economic organization (5)
- Leadership and organizational behaviour (6)
- Marketing (7)
- Strategy and logistics (8)

17. Indicate your academic status

- Professor (1)
- Associate Professor (2)
- Researcher (3)
- Senior Lecturer (4)
- Lecturer (5)
- Postdoctoral Fellow (6)
- Ph.D Student (7)
- Other (8)

Please provide any additional comments regarding research data management, potential data curation, or this survey

Appendiks D. Årsaker til bevaring av data

Fullstendig oversikt over svar i spørsmål 8 «Please explain why you plan on keeping the data for this amount of time»

	Fremsatte argument
Gjenbruk av data	<p>Ofte kan data brukes til flere prosjekter.</p> <p>It would be interesting to have more surveys and see how the dynamics change through time.</p> <p>use again until too old</p> <p>In case further interesting analyses can be made, e.g., if the data could be connected with other data sets in the future.</p> <p>Rich data sources may be used for analysis of several research questions.</p> <p>Raw data can be useful for other researchers or for myself in my future researches. It is always good to have (as much as possible) not-biased raw data (i.e. data that can be interpreted in another way, on the basis of a different theory, a different point of view, etc.)</p> <p>Data gathered from interviews, documents, etc are likely to be useful over again.</p> <p>May use it for longitudinal analysis</p> <p>you just don't delete the data you have collected, perhaps you find a use for it in a new project (this does not, of course, apply to data you contractually have to delete, e.g. from the SSB)</p> <p>historical analysis, time series</p> <p>valuable data which was gathered also manually from records, can be used for new projects</p> <p>The data I use is all publicly available. The main reason for me to keep it myself is in case I want to reuse it in future projects. The total size is also quite small, so keeping it is basically costless.</p> <p>Following up research</p> <p>Multiple projects that may span from the data.</p> <p>I want to make longitudinal comparisons. I also want to be able to refer back to my data if there are questions.</p> <p>in case some new idea comes up.</p> <p>Hate to throw things away. The data could be useful in the future</p> <p>Analyze changes over time. Data owner has to approve prologation of the project</p> <p>For two reasons: (1) most projects take a long time to complete, 2 - 12 years depending on factors such as how cooperation with coauthors progress, review time etc. for example I've just published a paper using data gathered in 2006. We submitted the first version of that paper to a journal in 2010. That version was rejected. We then sent the paper to another journal in 2012. This journal required us to do two major revisions and two minor ones before accepting the paper in 2015. We started writing another paper using these data in 2008, however the paper is not yet published because we have prioritized working on other papers. (2) I plan on storing my datasets for an indefinite time. I never know when that dataset might become useful or when I might get new research ideas that might utilize those data. Also, there's really no good arguments that I should delete a dataset. Datasets are costly to acquire, why delete them?</p>
Etterprøvarhet	<p>Kan brukes i andre studier - for å dokumentere tidligere studier</p> <p>Journal requirement and allowing replication</p> <p>There is an increased interest in sharing data, asking for renewed submissions in case of doubts and protests. All practices indicate that researchers will be increasingly accountable for documenting the veracity of their data in the years to come.</p> <p>Inquiries from other researchers, possibility of revisiting the topic and comparing old data with new.</p>

Required by journals of my field

Journals usually require this

Allow other researcher to replicate and build on my research. Also easy for myself to reproduce my own research.

In the case I need them again or I need to replicate my results

Just in case someone wants to review or retest my data, but if nothing has happened after 6 years I assume noone ever will, so why keep it? I do, however see the poitin in keeping it infinitely, but I don't do that as of now. If so, I need a much better personal storage system.

publications

need to answer questions from other researchers

Until interest in the work has subdued, any disagreement has surfaced, and I would no longer be required to defend findings.

Documentation, required by several scientific journals, in case of need for new analyses.

For research integrity, the data should be available for other researchers/academic journal; and for potential future project

For replication purposes

That's what journals want.

In case journal asks for a review of results.

Lagringslengde

Forefront research should be actual for many years

Data should, in principle, be kept indefinitely.

That is the amount of time my PhD project takes, in addition to publication time.

I have a limited contribution on the project and the field of the project is not within my narrow research interests.

I am a slow person.....

To have time to publish from the data.

The data is in the CCGR data basem which serves every reseracher at BI . We add data to it now and then

Why not

No specific reason

Has value

I do not need the data after this period.

No long time interest

Contracts with Statistics Norway and Centre of Corporate Governance Research (CCGR) at BI specify (tentative) dates of deletion of research data, often agreed on from our expected publication dates.

Sometimes it is not allowed to keep the data longer. But I would be happy to keep the data for longer time since one can have many ideas by using the same data and it takes time from starting the papers to getting them published.

Appendiks E. Kommentarer om tema forskningsdata eller undersøkelsen

Spørsmål 21. "Please provide any additional comments regarding research data management, potential data curation, or this survey"

I think training in research data management should be really beneficial.

Jeg tror dette er en viktig ting vi må begynne å ta alvorlig. Godt initiativ, håper vi får skikkelig ordning på dette!

For me this seems to be an interesting question regarding the type of data I use. Firstly, there are anonymity issues, both regarding the individual and the organization. Secondly, with top scientific journals requiring that data only be used limitedly, this applied to both qualitative and quantitative sets, I do not see the purpose of making the data available.

No initiation is really necessary. All good researchers are already taking care of their data. If any journal or other researcher needs our data, they can just email us.

BI should establish infrastructure which makes it as easy as possible for researchers to use individual level register data (e.g. from Statistics Norway).

Since I am the only postdoctoral fellow in my department, I did not experience that my anonymity was well attended to.

this is a non-issue

Maybe you should have asked before completing the questionnaire, whether I am doing Research at BI. I am in a part time position and is not expected to do Research at BI

The biggest data problem we face in my department, is the cost and (slow) speed of getting data from the SSB which does not seem to be geared towards supply data to researchers outside the SSB. How can BI help with this? It can help put pressure on politicians and managers in the SSB. Right now the Frisch center and the SSB has access to lots of data, but the rest of us are excluded.

PhD students often struggle to find organisations interested to serve as data samples. The access is uneven, depending highly on projects and connections of one's supervisor. Is there any way to make the access more evenly distributed, so that all PhD student who need observational data can be guaranteed at least some sort of "minimum samples"? A guarantee would then be a part of the PhD contract with BI.

The most important contribution to enhancing BI 's research output would be to have maximum access to interesting databases and to have capacity for statistical analysis available for researchers. Any serious medical research institution has statistical support for the medical researchers. Business schools rely shamelessly on amateurs. Let us get up to speed!!

it would be nice to have more databases which can show the information not only on the particular period in time, but also over time. Some of these options are available in such useful database as Orbis (ownership data), but BI is not subscribed to the data over time. There are also several valuable modules in Compustat, which BI is not subscribed.

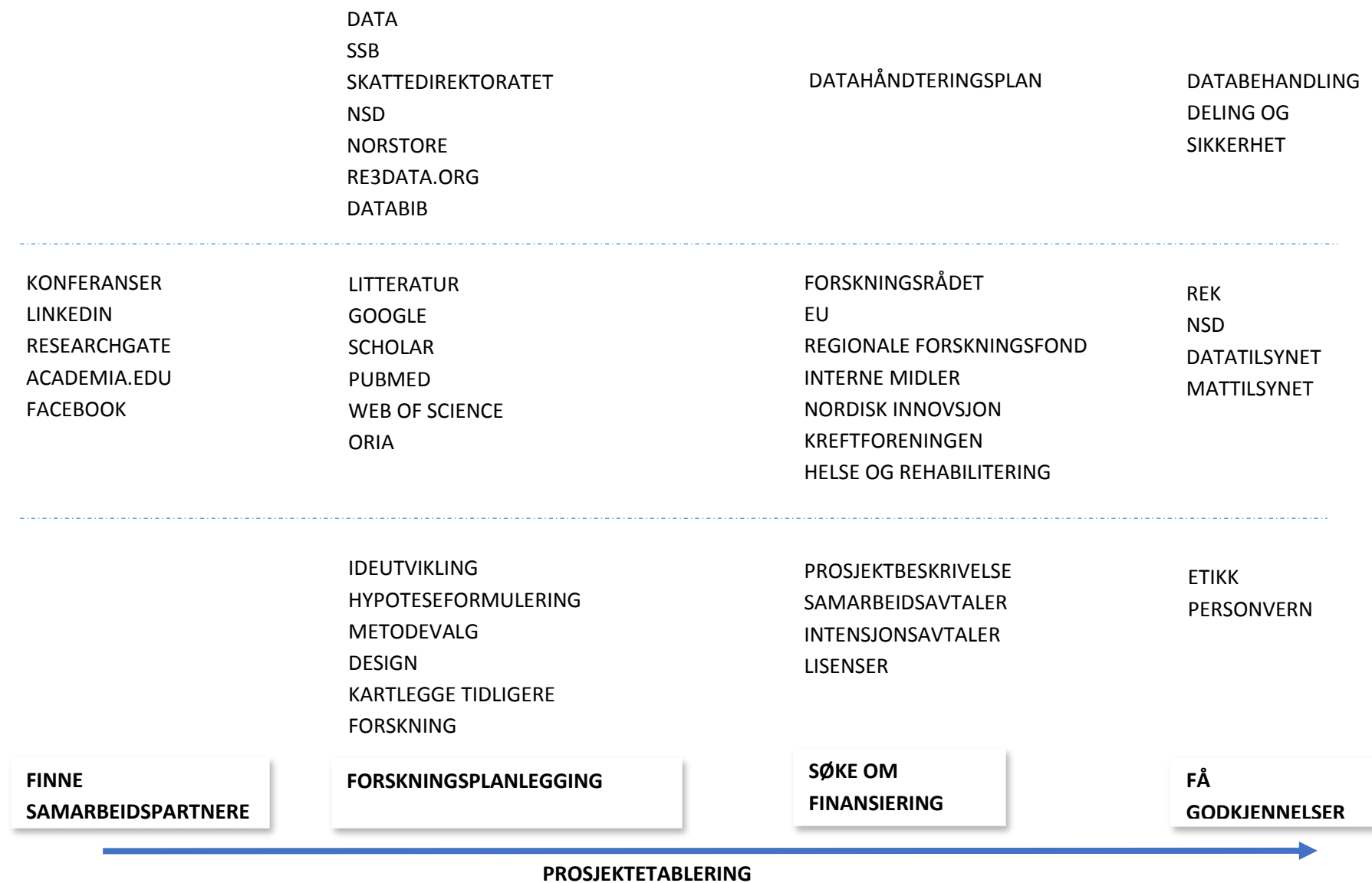
When BI is willing to subscribe to new data sets among which it has to choose a limited number, a survey can be sent to all the scholars to evaluate the necessity of the data set.

No

Good initiative !

Researchers at BI have faced some challenges with getting hold of data from Statistics Norway, including long waiting times, high prices etc. As far as I understand, there is now some work towards centralizing and coordinating these kinds of data requests. One could imagine these kinds of data sets being part of a common data infrastructure/pool at BI where e.g. new researchers/PhD students could get access.

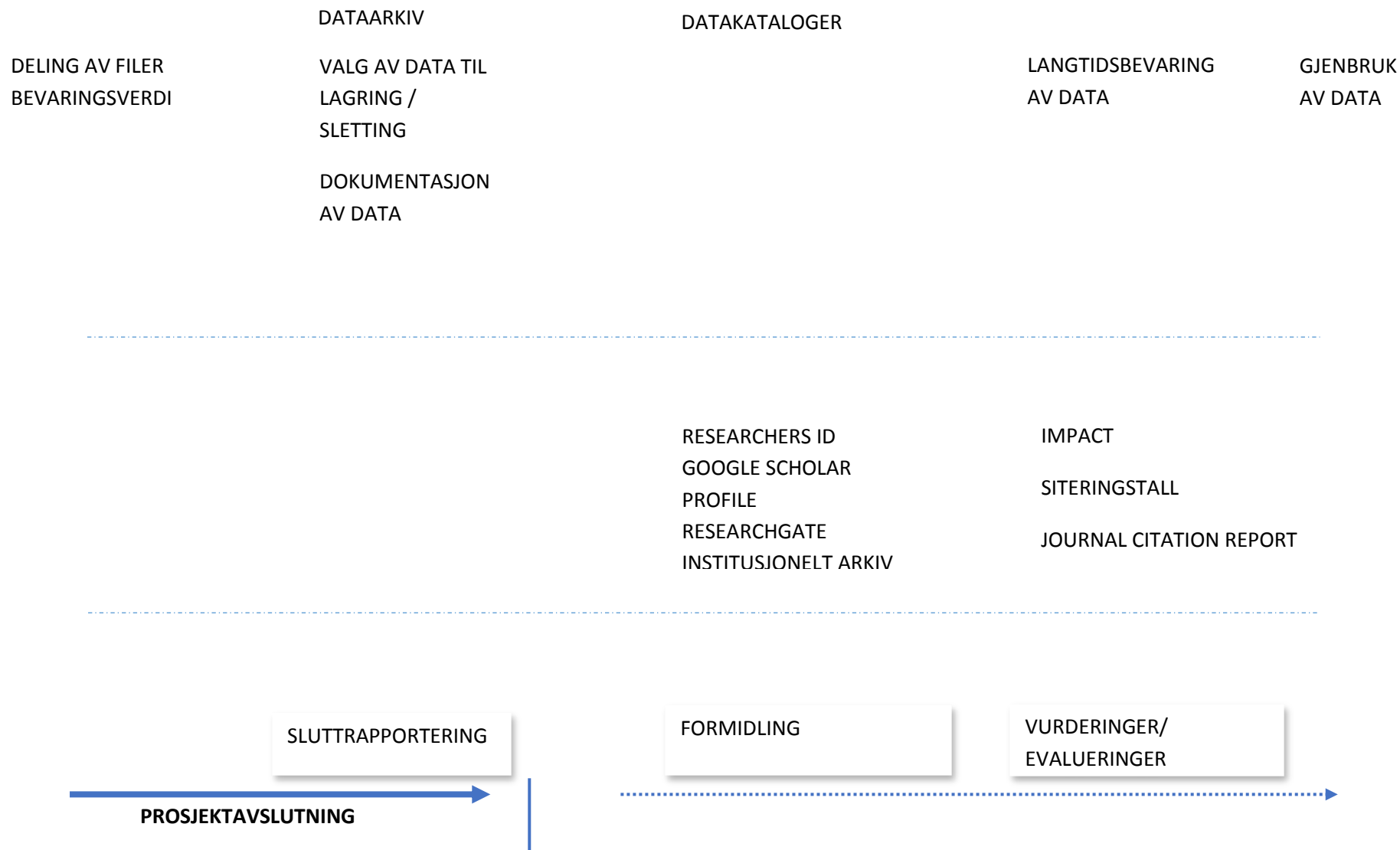
Appendiks F. Forskningsdatas gang i forskningsprosessen



Figur 5. Forskningsdata ved prosjektetablering



Figur 6. Forskningsdata ved prosjektgjennomføring



Figur 7. Forskningsdata ved prosjektavslutning

Referanser

- Andreoli-Versbach, Patrick, & Mueller-Langer, Frank. (2014). Open access to data: An ideal professed but not practised. *Research Policy*, 43(9), 1621-1633. doi: <http://dx.doi.org/10.1016/j.respol.2014.04.008>
- Aukland, Mary. (2012). *Re-skilling for Research: An investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers* Lastet ned fra <http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf>
- Bell, Gordon, Hey, Tony, & Szalay, Alex. (2009). Beyond the Data Deluge. *Science*, 323(5919), 1297-1298. doi: 10.1126/science.1170411
- Borgman, Christine L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge: MIT Press.
- Braman, Sandra. (2006). *Change of state: information, policy, and power*. Cambridge, Mass.: MIT Press.
- Corrall, Sheila, Kennan, Mary Anne, & Afzal, Waseem. (2013). Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research. *Library Trends*, 61(3), 636-674. doi: 10.1353/lib.2013.0005
- Cox, Andrew M. (2012). An exploration of the practice approach and its place in information science. *Journal of Information Science*, 38(2), 176-188. doi: 10.1177/0165551511435881
- Cox, Andrew M., & Pinfield, Stephen. (2013). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*. doi: 10.1177/0961000613492542
- Data Asset Framework. (2009). *Data Asset Framework Implementation Guide* Lastet ned fra http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf
- Digital Curation Center. (2013). *Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO): Workflow Planning and Documentation* Lastet ned fra http://www.dcc.ac.uk/sites/default/files/documents/Cardio_Characteristics_By_Statement.pdf
- Digital Curation Center. (2015). How to Guides & Checklists. Lastet ned 26.05., 2015, fra <http://www.dcc.ac.uk/resources/how-guides>
- Dillo, Ingrid, & De Leeuw, Lisa. (2014). *Data seal of approval: Certification for sustainable and trusted data repositories (2014-2015)* Lastet ned fra http://datasealofapproval.org/media/filer_public/2013/09/27/guidelines_2014-2015.pdf
- European Commission. (2013). *Guidelines on Data Management in Horizon 2020: The EU Framework Programme for Research and Innovation*. Lastet ned fra http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Fink, Arlene. (2013). *How to conduct surveys: a step-by-step guide*. Los Angeles: Sage.
- Gilje, Nils, & Grimen, Harald. (1995). *Samfunnsvitenskapenes forutsetninger: innføring i samfunnsvitenskapenes vitenskapsfilosofi*. Oslo: Universitetsforlaget.
- Helseforskningsloven. (2008). Lov om medisinsk og helsefaglig forskning (helseforskningsloven). Lov av 20. juni 2008 nr. 44: <http://lovdata.no/dokument/NL/lov/2008-06-20-44?q=helseforskningsloven>
- Helseregisterloven. (2001). Lov om helseregistre og behandling av helseopplysninger (helseregisterloven). Lov av 18. mai 2001 nr. 24, sist endret ved lov av 21. juni 2013 nr. 08: <http://lovdata.no/dokument/NL/lov/2001-05-18-24?q=helseregisterloven>.
- Jones, Sarah, Pryor, G., & Whyte, A. (2013). *How to develop RDM Services – a guide for HEIs* Lastet ned fra http://www.dcc.ac.uk/sites/default/files/documents/publications/How-to-develop-RDM-services_finalMay2013rev.pdf

- Jones, Sarah, Ross, Seamus, & Ruusalepp, Raivo. (2009). *Data Asset Framework Methodology* Humanities Advanced Technology and information Institute (HATII) (Red.), Lastet ned fra http://www.data-audit.eu/DAF_Methodology.pdf
- Kruse, Filip, & Thestrup, Jesper B. (2014). Research Libraries' New Role in Research Data Management, Current Trends and Visions in Denmark. *LIBER Quarterly*, 23(4), 301-335. <https://liber.library.uu.nl/index.php/lq/article/view/9173/9922>
- Kuipers, Tom, Hoeven, Jeffrey van der, & Parse Insight. (2009). *Insight into digital preservation of research output in Europe* Lastet ned fra http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Kunnskapsdepartementet. (2009). *Klima for forskning* . Hentet fra: <http://www.regjeringen.no/pages/2178785/PDFS/STM200820090030000DDDPDFS.pdf> (Vol. 30 (2008-2009)). Oslo: Departementenes servicesenter.
- Kunnskapsdepartementet. (2013). *Lange linjer - kunnskap gir muligheter*, hentet fra: <http://www.regjeringen.no/nb/dep/kd/dok/regpubl/stmeld/2012-2013/meld-st-18-20122013.html?id=716040> (Vol. 18(2012-2013)). Oslo: Departementenes servicesenter, Informasjonsforvaltning.
- Kunnskapsdepartementet. (2014). *Tilstandsrapport: Høyere utdanning 2014* Lastet ned fra https://www.regjeringen.no/globalassets/upload/kd/vedlegg/rapporter/tilstandsrapport2014_endelig_versjon.pdf
- Kvale, Live Håndlykken. (2012). *Data Sharing in the Life Sciences: A Study of Researchers at The Norwegian University of Life Sciences*. Live H. Kvale, Oslo. Lastet ned fra <http://hdl.handle.net/10642/1269>
- Kvale, Steinar. (1999). Forskere i lære. I S. Kvale & K. Nielsen (Red.), *Mesterlære: læring som sosial praksis* (s. 149-163). Oslo: Ad Notam Gyldendal.
- Larsen, Asger Væring, Zurcher, Sacha, Andersen, Mikkel Hvidtfeldt, Thestrup, Jesper Boserup, Mikkelsen, Søren Ærendahl, Sandfær, Mogens, & Conrad, Anders. (2014). *FIF - Fælles Infrastruktur for Forskningsdata* Lastet ned fra http://www.deff.dk/fileadmin/user_upload/dokumenter/DEFF/aktuelt/Afrapporteringer/FIF_afsluttende_rapport_20140718.pdf
- Laursen, Liv Bjerger, Brøndum, Iben, Davidsen, Katrine Düring, Hagelskjær, Anne, & Svensson, Ditte Schjødt. (2013). *Forskertjenester i danske FFU-biblioteker: best practice // next practice* *Netværk for Forskningsunderstøttende Services - NEFUS-rapport* Lastet ned fra http://dfdf.dk/index.php/rapporter/doc_download/151-rapport-om-forskertjeneste-i-danske-forskningsbiblioteker
- Lave, Jean, & Wenger, Etienne. (1991). *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.
- League of European Research Universities. (2013). Roadmap for Research data. LERU Research Data Working Group *Advice paper No.14. December 2013*. http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf
- McGowan, T., & Gibbs, T. A. . (2009). *Southampton Data Survey: Our Experiences & Lessons Learned* [unpublished]. . UK: University of Southampton.
- Merton, Robert K. (1957). Priorities in scientific discovery: a chapter in the sociology of science. *American sociological review*, 635-659. <http://www.jstor.org/stable/2089193>
- Mitroff, Ian I. (1974). Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. *American Sociological Review*, 579-595. <http://www.jstor.org/stable/2094423>
- National Science Board. (2005). *Long-Lived Digital Data Collection* Lastet ned fra <http://www.nsf.gov/geo/geo-data-policies/nsb-0540-1.pdf>
- Nielsen, Hans Jørn. (2012). Videnskabelig kommunikasjon og forskningsbiblioteker i en digital tid. I H. Høystrup, H. J. Nielsen & B. Hjørland (Red.), *Viden i spil: forskningsbibliotekernes funksjoner i forandring* (s. 67-92). Frederiksberg: Samfundslitteratur.

- Norges forskningsråd. (2008). Rapportering til KD: OECDs Prinsipper og retningslinjer for tilgang til offentlig finansierte data. Oslo: Norges forskningsråd.
- Norges forskningsråd. (2012a). *Verktøy for forskning - Del I: Nasjonal strategi for forskningsinfrastruktur 2012-2017*. Hentet fra: http://www.forskningsradet.no/prognett-infrastruktur/Norsk_veikart_for_forskningsinfrastruktur/1253972136981
- Norges forskningsråd. (2012b). *Verktøy for forskning del II : Norsk veikart for forskningsinfrastruktur 2012-2017*. Hentet fra: http://www.forskningsradet.no/prognett-infrastruktur/Norsk_veikart_for_forskningsinfrastruktur/1253972136981
- Norges forskningsråd. (2013). *Store satsinger 2015* Lastet ned fra <http://www.forskningsradet.no/servlet/Satellite?blobcol=urldata&blobheader=application%2Fpdf&blobheadername1=Content-Disposition%3A&blobheadervalue1=+attachment%3B+filename%3D%229788212032491.pdf%22&blobkey=id&blobtable=MungoBlobs&blobwhere=1274503521901&ssbinary=true>
- Norges forskningsråd. (2014a). Sharing research data 2014 DAMVAD-NFR. Endelig rapport mai 2014. Lastet ned 30.05., 2014, fra http://www.forskningsradet.no/no/Nyheter/Forskere_flest_positive_til_a_dele_forskningsdata/1253996263723/p1174467583739?WT.ac=forside_nyhet
- Norges forskningsråd. (2014b). *Tilgjengeliggjøring av forskningsdata* Lastet ned fra http://www.forskningsradet.no/no/Nyheter/Forskningsdata_skal_deles/1254000298821/p1174467583739
- Organisation for Economic Co-operation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding* Lastet ned fra <http://www.regjeringen.no/upload/KD/Vedlegg/Forskning/OECD%20Principles%20and%20Guidelines%20for%20Access%20to%20Research%20Data%20from%20Public%20Funding.pdf>
- PARADE. (2009). *Strategy for a European Data Infrastructure* Kimmo Koski, Claudio Gheller, Stefan Heinzl, Alison Kennedy, Achim Streit & Peter Wittenburg (Red.), Lastet ned fra <http://www.csc.fi/english/pages/parade>
- Parham, Susan Wells. (2010). Testing the DAF for Implementation at Georgia Tech. <https://smartech.gatech.edu/handle/1853/39786>
- Parham, Susan Wells, Bodnar, Jon, & Fuchs, Sara. (2012). Supporting tomorrow's research Assessing faculty data curation needs at Georgia Tech. *College & Research Libraries News*, 73(1), 10-13. <http://hdl.handle.net/1853/48706>
- Personopplysningsloven. (2000). *Lov om behandling av personopplysninger (Personopplysningsloven)*. Lov av 14. april 2000 nr. 31. Sist endret ved Lov av 11. januar 2013 nr 3.: <http://lovdata.no/dokument/NL/lov/2000-04-14-31?q=personopplysningsloven>.
- Pilerot, Ola. (2013). A practice theoretical exploration of information sharing and trust in a dispersed community of design scholars. *Information Research*, 18(4), 18-14. <http://www.informationr.net/ir/18-4/paper595.html#.VTEEcY4VaQ>
- Porat, M.U. (1977). The Information Economy: Definition and Measurement. *Office of Telecommunications (DOC), Washington, DC*. <http://files.eric.ed.gov/fulltext/ED142205.pdf>
- Qualtrics. (2015). Qualtrics Survey Software. Version 60854. <http://www.qualtrics.com>. Provo, Utah, USA.
- RAIRD. (2015). RAIRD - Remote Access Infrastructure for Register Data. Lastet ned 26.05, 2015, fra <http://raid.no/>
- Reilly, S. , Schaller, W. , Schrimpf, S. , Smit, E. , & Wilkinson, M. (2011). Integration of Data and Publication; ODE report: Opportunities in Data Exchange, hentet fra: http://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf.
- Ringdal, Kristen. (2013). *Enhet og mangfold: samfunnsvitenskapelig forskning og kvantitativ metode*. Bergen: Fagbokforlaget.
- Rolando, Lizzy, Doty, Chris, Hagenmaier, Wendy, Valk, Alison, & Parham, Susan Wells. (2013). *Institutional Readiness for Data Stewardship: Findings and Recommendations from the Research Data Assessment* Lastet ned fra <http://hdl.handle.net/1853/48188>

- Skagen, Therese. (2015). *Lagring av forskningsdata i Norge - Hva tilbyr de største arkivene? Prosjektrapport*. Lastet ned fra <http://hdl.handle.net/11250/283600>
- Statistikkloven. (1989). Lov om offisiell statistikk og Statistisk Sentralbyrå (statistikkloven). Lov av 16. juni 1989 nr. 54, sist oppdatert ved lov av 11. januar 2013 nr. 3: <https://lovdata.no/dokument/NL/lov/1989-06-16-54>.
- Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research, 36*(2), 84-90. doi: 10.1016/j.lisr.2013.11.003
- Tenopir, Carol, Allard, Suzie, Douglass, Kimberly, Aydinoglu, Arsev Umur, Wu, Lei, Read, Eleanor, . . . Frame, Mike. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE, 6*(6), e21101. doi: 10.1371/journal.pone.0021101
- Thestrup, Jesper B., Kruse, Filip, Nondal, Lars, Dorch, Bertil F., Andersen, Mikkel, Blaabjerg, Niels Jørgen, . . . Knudsen, Ellen V. (2012). *Forvaltning af forskningsdata i Danmark: En undersøgelse af danske universiteters og forskningsråds praksisser, politik og visioner for lagring, langtidsbevaring, tilgængeliggørelse og deling af forskningsdata*. Hentet fra: <http://www.deff.dk/aktuelt/artikel/forvaltning-af-forskningsdata/>
- Thune, Taran Mari, Aamodt, Per O, & Gulbrandsen, Magnus. (2014). *Noder i kunnskapsnettverket: Forskning, kunnskapsoverføring og eksternt samarbeid blant vitenskapelig ansatte i UH-sektoren. NIFU-rapport 23-2014*. Lastet ned fra <http://www.nifu.no/publications/1152682/>
- Universitetet i Oslo. (2015). *Dataeksplosjonen - en utfordring, og en gedigen mulighet! Rapport fra arbeidsgruppen «Lagring og deling av forskningsdata» ved Universitetet i Oslo (UiO) 11.05.2015*
- Van den Eynden, V., & Bishop, L. (2014). *Incentives and motivations for sharing research data, a researcher's perspective. A Knowledge Exchange Report*. Lastet ned fra http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf
- Van der Graaf, M. , & Waaijers, L. (2011). *A Surfboard for Riding the Wave. Towards a four country action programme on research data. A Knowledge Exchange Report*, hentet fra www.knowledge-exchange.info/surfboard.
- Vlaeminck, Sven (2012). *Research Data Management in Economic Journals*. Hentet fra <http://openeconomics.net/resources/data-policies-of-economic-journals/>
- Webster, Frank. (2006). *Theories of the information society*. London: Routledge.
- Whyte, Angus. (2014). *A pathway to sustainable research data services: from scoping to sustainability*. IG. Pryor, S. Jones & A. Whyte (Red.), *Delivering research data management services: Fundamentals of good practice* (s. XIV, 242 s. : ill.). London: Facet Publishing.
- Whyte, Angus, & Tedds, Jonathan. (2011). *Making the Case for Research Data Management*. DCC Briefing Papers. Lastet ned 26.05., 2015, fra <http://www.dcc.ac.uk/sites/default/files/documents/publications/Making%20the%20case.pdf>
- Wood, John, Andersson, Thomas, Bachem, Achim, Best, Christoph, Genova, Françoise, Lopez, Diego R., . . . Hudson, Richard L. (2010). *Riding the wave : How Europe can gain from the rising tide of scientific data*. Hentet fra: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>: European Union.
- Østbye, Helge, Helland, Knut, Knapskog, Karl, & Larsen, Leif Ove. (2013). *Metodebok for mediefag*. Bergen: Fagbokforlaget.