

Dataartikelns roll i vetenskaplig kommunikation

En intervjustudie med forskare vid svenska universitet

Karin Sandqvist

Examensarbete (30 högskolepoäng) i biblioteks- och informationsvetenskap
för masterexamen inom ABM-masterprogrammet vid Lunds universitet.

Handledare: Jutta Haider

År: 2018

Title

The Function of Data Papers in Scholarly Communication:
An Interview-based Study of Researchers at Swedish Universities

Abstract

To make not only scholarly publications but also the data underlying those publications openly accessible is part of the movement for Open Science. There are several arguments for sharing research data, for example to promote science and collaboration, to make validation of results possible and to reduce redundant data production (Callaghan et al. 2012; Tenopir et al. 2011). Open Data policies are a more and more common strategy from governments, research funders as well as universities. Still – a lot of data is never shared and several studies have pointed out a number of reasons for this: Researchers don't feel they get rewarded for making data useful for others, they lack time and they fear that their research might get scooped. There are also concerns around sensitive data and intellectual property (Kratz & Strasser 2015; Tenopir et al. 2011). Sometimes researchers just don't consider their data to be relevant for others (Tam 2016).

The data paper is "a journal publication whose primary purpose is to describe data, rather than to report a research investigation. [...] Its purposes are threefold: to provide a citable journal publication that brings scholarly credit to data publishers; to describe the data in a structured human-readable form; and to bring the existence of the data to the attention of the scholarly community" (Chavan & Penev 2011, s. 3). While the reasons for *not* sharing data have been studied thoroughly (e.g. Van den Eynden & Bishop 2014) less is known about the incentives to actually share data, and even less about what role the data paper can play. Therefore the aim of this study is to deepen the understanding of what functions a data paper can have for researchers data sharing practices, how it differs from other ways of sharing data and what significance it has that a data paper can be reviewed, cited and part of the academic merit system. The theoretical framework is based on three different perspectives: Epistemical cultures (Knorr Cetina 1999), the socio-technical perspective (Bijker 1995; Van House 2003a) and the document perspective (Frohmann 2004). Qualitative interviews were performed with eight researchers who work in different research areas and in different positions at Swedish universities. All of them have written and/or cited data papers or otherwise actively shared their data.

Some of my findings are that: writing a data paper can be a faster way of publishing data, since you don't have to write a conventional article with hypotheses and argumentation; in a data paper you can publish null and negative findings; the data paper makes it easier to cite data in a formalized way; and it saves space in the applied article when you can describe data in detail in a data paper. Also the interviewed researchers see no risk that their openness will be misused and the fact that a data paper can be added to your cv is regarded as something that can be beneficiary for both your career and possibilities of receiving funding. The review process can enhance the data and gives it a quality mark.

Master's thesis

Keywords

Library and information science; scholarly communication; research; open data; open science; data sharing practices; data papers; qualitative research design; interviews.

*Stort tack till de forskare som tog sig tid att svara på mina frågor
och till min handledare Jutta Haider för inspiration och kloka synpunkter.
Tack också till Helena Francke, Jonas Fransson, Maria Johnsson,
Sara Kjellberg, Monica Lassi och Fredrik Åström för värdefulla uppslag.*

INNEHÅLLSFÖRTECKNING

1. INLEDNING.....	6
1.1 PROBLEMFORMULERING	7
1.2 SYFTE OCH FRÅGESTÄLLNINGAR	8
1.3 UPPSATSEN I RELATION TILL INFORMATIONSVETENSKAPLIG FORSKNING.....	8
1.4 ORDLISTA OCH VIKTIGA BEGREPP	10
1.5 DISPOSITION.....	11
2. FORSKNINGSDATA SOM EN DEL AV VETENSKAPLIG KOMMUNIKATION	12
2.1 FORSKNINGSDATA	12
2.1.1 Från stora data till små i den långa svansen	13
2.1.2 Data i relation till tid	14
2.1.3 Kvalitativa och kvantitativa data	15
2.2 DELNING AV FORSKNINGSDATA.....	18
2.2.1 Policyer och stöd för forskningsdatahantering.....	19
2.2.2 Etik och integritet	21
2.3 PUBLICERING AV DATA	22
2.3.1 Olika sätt att publicera data och publiceringens betydelse.....	22
2.3.2 Datatidskrifter och dataartiklar.....	23
3. TEORETISKA PERSPEKTIV	26
3.1 SOCIALA PRAKTIKER OCH EPISTEMISKA KULTURER	26
3.2 DATAARTIKELN SOM SOCIOTEKNISK ENSEMBLE	27
3.3 DOKUMENTETS ROLL.....	29
3.4 SAMMANFATTNING	29
4. INTERVJU SOM METOD.....	31
4.1 INTERVJUER.....	31
4.1.1 Urval.....	31
4.1.2 Genomförande av intervjuerna	32
4.1.3 Forskningsetiska överväganden	33
4.2 PRESENTATION AV RESULTATET OCH ANALYSENS GENOMFÖRANDE.....	34
4.3 EGEN HANTERING AV FORSKNINGSDATA	35
5. INTERVJUERNA MED FORSKARNA	38
5.1 VEM DELAR DATA – OCH MED VILKA ANVÄNDARE I ÅTANKE?	38
5.2 VILKA DATA DELAS OCH VILKA DELAS INTE?	40
5.3 NÄR DELAS DATA?.....	42
5.4 VARFÖR DELAS DATA?.....	43
5.5 HUR GÅR DET TILL ATT DELA DATA?	45
5.6 VAD INNEBÄR DET ATT DELA DATA?	47
6. VIKTIGA FRÅGOR VID DELNING OCH PUBLICERING AV DATA	50
6.1 FÖRTROENDE OCH TROVÄRDIGHET	50
6.2 SAMARBETE OCH ERKÄNNANDE	52
6.3 PÅVERKAN OCH STATUS	54
6.4 SAMMANFATTNING	56
7. DISKUSSION OCH SLUTSATSER	57
7.1 ÅTERKOPPLING TILL FORSKNINGSFRÅGORNA	57
7.2 AVSLUTANDE DISKUSSION	60
7.2.1 Framtida forskningsfrågor.....	61
7.2.2 Delade data för fortsatta upptäckter	62
KÄLLFÖRTECKNING	63

<i>Bilaga 1 · Intervjuförfrågningar</i>	<i>69</i>
<i>Bilaga 2 · Intervjuguide</i>	<i>70</i>
<i>Bilaga 3 · Informerat samtycke</i>	<i>71</i>
<i>Appendix · Forskarna och intervjuerna</i>	<i>73</i>

1. Inledning

När astronomer i början av 2018 upptäckte att det troligen finns tusentals svarta hål i centrum av vår galax var det inte – som man lätt skulle kunna tro – för att de hade fått tillgång till ny spjutspetsteknologi i ett superteleskop (Meyer 2018). I Smithsonian Magazine beskrivs hur denna upptäckt istället hade gjorts genom att forskare grävde fram arkiverade data, varav några var nästan tjugo år gamla. Skribenten Eileen Meyer, själv astronom, konstaterar att sådana här upptäckter kommer att bli allt vanligare eftersom ”the era of ’big data’ changes how science is done” (ibid. 2018). Dagens astronomer samlar in en ständigt ökande mängd data som kommer att möjliggöra nya upptäckter lång tid framöver. Om sin egen forskning, där hon använt fyrahundra arkivbilder från en tidsperiod av tretton år, skriver hon:

This kind of work was only possible because other observers, for other purposes, just happened to capture images of the source I was interested in, back when I was in kindergarten. As astronomical images become larger, higher resolution and ever more sensitive, this kind of research will become the norm.

Meyer 2018¹

Detta är bara ett av många exempel på hur digitala och tekniska framsteg förändrar vetenskapens villkor och möjliggör dels forskning på stora mängder data, dels forskning som i sin tur genererar allt större mängder data (Borgman 2007). Det har också blivit enklare att dela data, en utveckling som har bidragit till att data inte längre ses endast som en biprodukt av enskilda forskningsprojekt utan som värdefulla i sig och något som kan inspirera till nya hypoteser och möjliggöra nya vetenskapliga insikter och upptäckter (Tenopir et al. 2011).

Men för att sådana framtida upptäckter ska vara möjliga måste inte bara själva forskningsresultaten, i form av publikationer, utan även de data som ligger till grund för resultaten göras möjliga att hitta och återanvända. Rörelsen för *Open access* till vetenskapliga artiklar (Suber 2012) har därför följts av diskussioner om *Open data* eller *Open Science*. Fördelarna sägs vara många: Forskningsresultat kan verifieras, nya samarbetsmöjligheter – särskilt tvärvetenskapliga – kan uppstå, överflödiga datainsamling undviks och forskningsdata kan användas på nytt, för jämförande studier och för helt ny forskning (Callaghan et al. 2012; Tenopir et al. 2011). På så sätt sparas resurser och forskningens ekonomiska, sociala och ekologiska hållbarhet ökar (Chowdhury 2014). Öppen vetenskap och öppna data har blivit ledord för såväl lärosäten, myndigheter, forskningsfinansiärer och vetenskapliga förlag som för många enskilda forskare – både nationellt och internationellt (Europakommissionen u.å.; Vetenskapsrådet 2015). I Sverige är målet att ”en omställning

¹ Webbartikel, sidhänvisning saknas.

till öppen tillgång till forskningsresultat inklusive vetenskapliga publikationer, konstnärliga verk och forskningsdata bör vara genomförd fullt ut senast inom tio år” (Utbildningsdepartementet 2016, s. 107).

I detta sammanhang är det intressant att studera dataartiklar (på engelska *data papers*), en relativt ny genre inom vetenskaplig kommunikation. Dataartiklar beskriver delade data samt var, varför, hur och av vem informationen har samlats in (Callaghan et al. 2012). På samma sätt som en konventionell, vetenskaplig artikel kan en dataartikel granskas, publiceras, citeras och utgöra ett mått för forskningens genomslag. Frågan om dataartiklars betydelse är viktig eftersom det är vanligt inom många forskningsområden att data inte delas annat än informellt och inom en liten krets (Akers 2016). En vanlig förklaring till detta är att det tar för mycket tid (Kratz & Strasser 2015). Till skillnad från vetenskapliga publikationer, som är dokument skapade för att kunna spridas och stå för sig själva, måste data nämligen sättas i sin kontext (Borgman 2015, s. 256, Mauthner & Parry 2013), något som är tidskrävande. Ändå är detta i regel mycket billigare än att samla in data på nytt – i de fall det ens är möjligt att rekonstruera insamlingsprocessen (Kratz & Strasser 2014). Ett problem i sammanhanget är att det ofta är samhället i stort eller andra forskare som tjänar på att data delas. Den forskare som lagt tid på att samla in och tillgängliggöra data har ofta fått betala med sin tid och sina forskningsmedel, utan att detta har avspeglats på cv:t eller värderats när forskningsmedel ska delas ut och tjänster tillsätts. Nya möjligheter och incitament för forskare att tillgängliggöra sina data skulle därför kunna bidra till den önskade övergången till öppen och hållbar vetenskap. Dataartiklar kan då möta både behovet att premiera och ge erkännande till forskare som delar sina data, och behovet de som vill återanvända data har av att få reda på hur data har tagits fram och granskats (Park & Wolfram 2017). Denna uppsats utgår från intervjuer med åtta forskare på svenska lärosäten som har skrivit och/eller citerat dataartiklar eller på annat sätt aktivt har valt att göra sina data öppet tillgängliga. Målet är att bidra till en fördjupad förståelse för dataartiklars betydelse i förhållande till utvecklingen av öppen vetenskap.

1.1 Problemformulering

Dataartiklar är en relativt ny genre inom vetenskaplig kommunikation som kan publiceras både som tillägg till en vetenskaplig artikel och helt fristående i renodlade datatidskrifter (på engelska *data journals*). Eftersom många datatidskrifter inte indexeras i bibliografiska databaser som Scopus och Web of Science det är det svårt att ta reda på exakt hur många som har publicerats. Open Data Report² (Elsevier & CWTS 2017) konstaterar dock att antalet citerade dataartiklar har vuxit fort och att deras betydelse ökar. Trots den snabba tillväxten av antalet datatidskrifter så känner många forskare inte till denna möjlighet att dela data (Akers 2016). Syftet med en dataartikel är att detaljerat beskriva ett datasets struktur och format, de metoder och protokoll som använts för att skapa och bearbeta datasetet samt möjliga sätt att återanvända det. Till formen liknar dataartikeln en traditionell vetenskaplig artikel, men den innehåller varken analyser, resultat eller de slutsatser som dragits utifrån beskrivna data (Candela, Castelli, Manghi & Tani 2015). Sådana analyser och slutsatser kan ibland publiceras i en applicerad artikel, men dataartikeln

² Rapporten gjordes i samarbete mellan Leiden University's Centre for Science and Technology Studies (CWTS) och det vetenskapliga förlaget Elsevier.

kan också vara helt fristående. Fördelarna för forskarna är att de får erkännande för sitt arbete med att tillgängliggöra data eftersom en dataartikel kan indexeras och citeras på samma sätt som en konventionell tidskriftsartikel. De kan på ett enkelt sätt följa vem som använder och citerar deras data, och publiceringen ökar synligheten, användbarheten och trovärdigheten för data (GBIF u. å.).

Det finns flera skäl att studera vad som driver forskare att dela sina data. Även om den allmänna nyttan med datadelning är stor så är det mer osäkert vad enskilda forskare tjänar på att dela sina data, som vi såg i föregående avsnitt. Detta gör det viktigt att undersöka normer och forskarnas drivkrafter (Van den Eynden & Bishop 2014). Ett stort antal studier har gjorts om vad som *förhindrar* delning av data (Borgman 2012; Van den Eynden & Bishop 2014). Mindre är känt om vilka *drivkrafter* som får forskare att dela och återanvända data och Christine Borgman (2012) betonar att mycket mer forskning behövs inom dessa områden. De undersökningar som har gjorts om vad som driver forskare att dela data är dessutom ofta kvantitativa i form av enkäter vilket gör det svårare att ställa följdfrågor och ta del av fördjupande resonemang. Inte heller har svenska förhållanden undersökts i någon större utsträckning. Resultaten i denna uppsats kan därför vara av intresse för personer som är involverade i arbetet med delning av egna eller andras forskningsdata, både rent konkret och på en mer övergripande nivå med exempelvis forskarstöd eller policyfrågor.

1.2 Syfte och frågeställningar

Syftet med studien är att skapa fördjupad förståelse för dataartiklars betydelse i forskningsprocessen. På så sätt hoppas jag kunna bidra till att utveckla kunskapen om öppna forskningsdata som en del av vetenskaplig kommunikation.

Studien utgår från följande forskningsfrågor:

1. Vilka funktioner kan en dataartikel fylla för forskare i deras arbete?
2. Vad skiljer en dataartikel från andra sätt att dela, publicera och återanvända data?
3. Vilken roll kan dataartiklar ha för att möjliggöra och underlätta återanvändning av data?
4. Vilken betydelse har det att en dataartikel kan granskas, citeras och vara en del av det akademiska meriteringssystemet på samma sätt som en konventionell vetenskaplig artikel?

1.3 Uppsatsen i relation till informationsvetenskaplig forskning

Forskningsdata som en del av vetenskaplig kommunikation kan, inom ramen för informationsvetenskaplig forskning, undersökas ur en rad aspekter. Att data inte ska ses som varje forskares enskilda egendom utan delas med forskarsamfundet och samhället i stort, är helt i enlighet med Robert K. Mertons principer och särskilt den om ”communalism”³

³ En av sociologen Robert K. Mertons så kallade CUDOS-normer (där CUDOS står för Communalism, Universalism, Disinterestedness och Organized Scepticism). Källa: Oxford Reference.

– att vetenskapliga framsteg och kunskap är kollektiva nyttigheter. Mertons principer har ibland kritiserats för att vara alltför funktionalistiska och för att förenkla vetenskapens mekanismer till ett självreglerande system av normer, värden och belöningar oberoende av sociala faktorer (Van House 2003a). Inte heller informationsteknologin hade någon central betydelse för Merton (Borgman 2007). Många studier av vetenskaplig kommunikation har i stället utgått från ett mer konstruktionistiskt perspektiv och har visat att forskares praktiker – snarare än att styras av normer – är sociala, lokalt situerade och varierande (Borgman 2007).

Ett sociotekniskt perspektiv, där sociala och tekniska faktorer antas vara lika viktiga och påverka varandra ömsesidigt (Van House 2003a), är vanligt när skärningspunkten mellan teknologi och beteende studeras. Enligt Ji-Hong Park (2008) kan sådan forskning delas in i två huvudområden där informationsvetenskaplig forskning och teknik- och vetenskapsstudier (STS⁴) överlappar varandra. Det ena området rör *processen* att sprida och använda vetenskaplig information, och det andra *strukturen* hos den kunskap som skapas när vetenskaplig information sprids och används. Park kopplar ihop process med konstruktivism och studier av exempelvis publicering, sökstrategier, samarbeten och skrivande med utgångspunkt i aktör-nätverk-teori, *Social Construction of Technology* (SCOT) eller laboratoriestudier. Struktur kopplar han ihop med Mertonsk funktionalism, utvärderande bibliometrisk studier, och forskning om relationer mellan forskningsområden och discipliner samt om innovationer och forskningsproduktivitet. Forskning om vetenskaplig kommunikation återfinns även inom andra discipliner som filosofi, sociologi, pedagogik samt media- och kommunikationsvetenskap (Park 2008).

Jag har valt att studera vilken betydelse dataartikeln kan ha för några av de forskare som har skrivit, granskat eller citerat dataartiklar. Därmed anknyter uppsatsen till tidigare informationsvetenskaplig forskning om hur *processer* inom vetenskaplig kommunikation utformas i samspel mellan människor, teknik och artefakter och hur dessa processer kan se olika ut i olika grupper.

En avgränsning har gjorts i det att jag inte går närmare in på forskningsbibliotekens roll. Inom ramen för arbetet med forskarstöd har biblioteken visserligen blivit allt mer involverade i arbetet med forskningsdata under olika stadier i forskningsprocessen. Tjänsterna kan vara både *rådgivande* och röra exempelvis metadatastandarder, datahanteringsplaner eller sökningar efter dataset, och *tekniska*, som i att tillhandahålla infrastruktur i form av nödvändiga program och repositorer eller att göra det möjligt att spåra och mäta forskningsdatas genomslag (Tenopir, Sandusky, Allard & Birch 2014). Detta arbete är sedan flera år under utveckling vid svenska universitet och lärosäten (se till exempel Anderberg 2016; Franssön et al. 2016; Johnsson & Lassi 2016). Forskningsbiblioteken kan också fungera som ett stöd när det gäller open access-publicering av konventionella vetenskapliga artiklar (se till exempel Bernhardt 2016; Robertson & Simser 2013). Det är troligt att även forskarstöd runt olika former av publicering av data kommer att bli en allt vanligare uppgift för biblioteken och därför är uppsatsens frågeställningar mycket relevanta för både biblioteksforskningen och professionen. Just när det gäller publicering av dataartiklar har jag dock inte funnit några exempel på att biblioteken har varit involverade i arbetet i någon större utsträckning och därför valde jag att göra denna avgränsning.

⁴ Efter engelskans Science and technology studies. STS uttolkas ibland också som Science, technology, and society, Social studies of science and technology (Van House 2003a) eller Sociotechnical systems (Borgman 2007).

I kapitel 2 görs en bredare forskningsöversikt där bakgrunden till uppsatsen och dess frågeställningar fördjupas genom att olika aspekter av forskningsdata som en del av vetenskaplig kommunikation belyses.

1.4 Ordlista och viktiga begrepp

APC (Article Processing Charge): En avgift många vetenskapliga förlag numera tar ut för att publicera en artikel. Avgiften ska täcka publikationskostnaderna och ses av många som ett sätt för förlagen att kompensera sig för förlorade prenumerationsintäkter när publicerade artiklar görs öppet tillgängliga. Avgiften betalas ofta av det universitet där forskaren är anställd eller av forskningsfinansiären.

Applicerad artikel: En av de forskare jag intervjuade använde uttrycket ”applicerad artikel” om artiklar där data som beskrivs detaljerat i en dataartikel används för att göra analyser och dra slutsatser. För att vara konsekvent använder jag själv uttrycket för att beskriva sådana artiklar som är kopplade till en dataartikel. Vanligen har då dataartikeln och den applicerade artikeln skrivits av samma person (antingen ensam eller tillsammans med en eller flera medförfattare). När jag jämför dataartiklar med ”vanliga” vetenskapliga artiklar använder jag istället uttrycket konventionell artikel.

Data och dataset: Sammanblandas ofta, men är, enligt Christine Borgman (2012), inte samma sak. Begreppet dataset kan vara praktiskt att referera till när det gäller en samling av data vid citering, men enligt Borgman hjälper termen inte till att förtydliga vad som menas med data. Problemet att urskilja lämpliga enheter av data att referera till är ett kärnproblem i försöken att etablera en praxis för datacitering. I föreliggande uppsats kommer begreppet dataset huvudsakligen att användas för att referera till en avgränsad enhet data i mer generella sammanhang.

Dataartikel: På engelska *data paper* (kallas ibland även *data descriptor*, *data note*, *data in brief article*, *data original article*, *data report* och *database article*). Beskriver en eller flera datasamlingar och möjliga sätt att återanvända dem, men innehåller normalt inte forskningsresultat, analyser eller slutsatser. En dataartikel bidrar till att öka möjligheten att finna och återanvända data och upphovspersonerna får erkännande för sitt arbete med att samla in och tillgängliggöra data genom att artikeln publiceras i en ofta peer reviewed tidskrift, kan sättas upp på ett forskar-cv och citeras som en konventionell artikel.

Datahanteringsplan: På engelska *Data Management Plan* (DMP). Ett formellt dokument som ger ett ramverk för vad som ska ske med datamaterialet under och efter forskningsprojektet. Källa: Svensk Nationell Datatjänst.

Datakurering: Att aktivt ta hand om data under hela dess livscykel, genom att exempelvis organisera, långtidsbevaka, beskriva, tvätta, ge mervärde till och främja användande av dessa data. Källa: Digital Curation Centre.

Digital Object Identifier (DOI): Ett permanent identifieringsnummer som används för att identifiera och lokalisera till exempel ett dataset eller en artikel i den digitala världen. Genom att ett dataset registreras och tilldelas ett DOI kan det hanteras som ett oberoende och unikt vetenskapligt objekt (Lawrence et al 2011).

Journal Impact Factor (JIF): Används som ett mått på tidskrifters vetenskapliga genomslag/synlighet utifrån hur ofta artiklar i tidskriften citeras under en viss tidsperiod.

Metadata: Metainformation eller data om data; beskrivning av innehållet i ett digitalt dokument. Metadata för texter, tabeller och webbsidor (i form av metataggar) kan beskriva vilket innehåll dokumentet har, dess språk och författare. Källa: Nationalencyklopedin.

Open access: Open access-litteratur är ”digital, online, free of charge, and free of most copyright and licensing restrictions” (Suber 2012, s. 4). *Gold OA* tillhandahålls av tidskrifter medan *green OA* tillhandahålls av repositorer (se nedan).

Repositorium: Ett digitalt arkiv för tillgängliggörande av texter eller data från forskare inom en viss institution, ett lärosäte eller inom en viss disciplin.

1.5 Disposition

I detta kapitel har uppsatsens forskningsproblem samt dess syfte och frågeställningar presenterats. Vidare har jag redovisat hur uppsatsen förhåller sig till tidigare forskning om vetenskaplig kommunikation ur ett biblioteks- och informationsvetenskapligt perspektiv och förklarat några viktiga begrepp. Kapitel två ger en bakgrund till arbetet med forskningsdata som en del av vetenskaplig kommunikation. Här beskrivs viktiga aspekter av forskningsdata, arbetet med delning och publicering av data samt forskningsdata i förhållande till öppen vetenskap, policyarbete och bibliometrisk forskningsutvärdering. Det teoretiska ramverket redovisas i kapitel tre. Uppsatsens empiriska material bygger på semistrukturerade intervjuer med forskare vid svenska lärosäten. Metodologiska överväganden, urval och genomförande av intervjuer och analys beskrivs i kapitel fyra. Resultaten av intervjuerna återfinns i kapitel fem och analyseras sedan i kapitel sex. I kapitel sju slutligen besvaras forskningsfrågorna och kapitlet innehåller även en diskussion om studiens betydelse och begränsningar och förslag på fortsatt forskning.

2. Forskningsdata som en del av vetenskaplig kommunikation

Forskning beskrivs vanligen som en cyklisk process, från forskningsidé och planering via datainsamling, tolkning och analys till publicering av forskningsresultat – som i sin tur leder till nya forskningsidéer. I samtliga faser samlar forskaren in, använder och/eller skapar nya data (Pryor 2012, s. 6). Den snabba tillväxten av mängden digitala data, internationaliseringen och det växande kravet på att offentligt finansierad forskning ska göras öppet tillgänglig för alla, har inneburit att tre nya faser har lagts till i denna forskningsprocess: bevarande, spridning och återanvändning av data (Kruse & Boserup Thestrup 2018, s. 3). Forskningsdata är därmed inte längre bara ett medel utan ses allt oftare som något som har ett värde i sig. Och samtidigt som forskningsområdets egenheter kan forma de sätt på vilka data hanteras, kan även själva forskningsområdet i sig omformas när data skapas, används och hanteras på nya sätt (Tam, Fry & Proberts 2014, s. 723). ”The challenges are to understand which data might be shared, by whom, with whom, under what conditions, why, and to what effects. Answers will inform data policy and practice,” skriver Christine Borgman i artikeln ”The Conundrum of Sharing Research Data” (2012, s. 1072).

I första kapitlet (se avsnitt [1.3](#)) satte jag uppsatsen i relation till några övergripande spår i den biblioteks- och informationsvetenskapliga forskningen om vetenskaplig kommunikation. Forsknings- och litteraturgenomgången i föreliggande kapitel är tänkt att ge en bakgrund till uppsatsens frågeställningar om dataartikelns betydelse genom en fördjupning i ämnet forskningsdata. Kapitlet är indelat i tre huvudavsnitt där det första tar upp olika centrala aspekter av forskningsdata. Det andra avsnittet beskriver aktiviteter runt och attityder till delning av forskningsdata som är viktiga i förhållande till öppen vetenskap. I det tredje avsnittet slutligen går jag närmare in på publicering av forskningsdata.

2.1 Forskningsdata

Frågan om vad som kännetecknar forskningsdata är nästan omöjlig att besvara eftersom data kan ha så många olika former – både fysiska och digitala. Fysiska forskningsdata kan utgöras av det mesta mellan himmel och jord – från stensamlingar och dinosaurieskelett till vävnadsprover och gener från bananflugor. Men även digitala data kan se ut på många olika sätt: De kan till exempel vara textbaserade, numeriska, relativa eller instrumentgenererade (Kennan 2017). Data kan också vara observationsbaserade eller i form av stillbilder, film, ljud, spel, och de kan vara insamlade eller skapade av såväl människor som maskiner och kräva olika instrument eller programvaror för att bli användbara (Uhlir & Cohen 2011 se Borgman 2015, s.19). Hur omfattande en lista över olika slags data än görs innebär den ändå inte en definition av data eftersom den inte klargör var gränsen går

mellan vad som inkluderas och inte i begreppet: "[A]ny such list is at best a starting point for what could be data to someone, for some purpose, at some point in time" (Borgman 2015). Det kan också vara svårt att skilja mellan vad som är forskningsdata och exempelvis arbetsmaterial, publikationer eller metadata. Här kan åsikterna skilja sig åt mellan olika forskningsområden och yrkesgrupper, men även från en forskare till en annan (Borgman 2015; Haider 2017). Det som är data för forskaren behöver alltså inte vara det för en bibliotekarie eller arkivarie – och omvänt kan det vara så att en forskare har en samling material utan att inse vilket värde det skulle kunna ha som data (Borgman 2012).

Alla försök att definiera vad forskningsdata är riskerar alltså att landa i så generella formuleringar att de blir mer eller mindre meningslösa. I stället vill jag i de tre följande avsnitten ta upp olika aspekter av forskningsdata som har betydelse i relation till öppen vetenskap: Det handlar om "stora data" och "små", data i relation till tid samt kvalitativa och kvantitativa data.

2.1.1 Från stora data till små i den långa svansen

Utvecklingen av ny teknik, och nya metoder att samarbeta, samla in och dela data har möjliggjort stora höginstrumentaliserade projekt med många inblandade forskare över hela världen. För att det ska vara möjligt att dela data i så stor omfattning krävs noggrant förarbete och dokumentation: "As more people are involved, and as time periods to design research and instrumentation become longer, greater consistency is required in the resulting data" (Wallis, Rolando, Borgman 2013, s. 3). Men så stora datamängder är bara vanliga inom några få forskningsfält, såsom astronomi, fysik, forskning om arvs massa, makroekonomi och delar av digitala humaniora (Borgman 2015, s. 9). Det mesta av all forskning sker i stället i mindre projekt, och även om varje enskilt projekt i sig sällan genererar mycket data blir den sammanlagda datamängden enorm eftersom det rör sig om så många forskningsprojekt (Heidorn 2008). Till skillnad från data i större projekt, som ofta bevaras och återanvänds, är risken stor att data från dessa mindre projekt går förlorade eftersom de inte indexeras och kureras på samma noggranna sätt. Bryan Heidorn (2008) kallar därför denna typ av data för mörka data ("dark data"). I artikeln "Shedding Light on the Dark Data in the Long Tail of Science" tillämpar han konceptet long-tail economics⁵ på forskningsdata: Även om det bara finns några få forskare i världen som behöver få tillgång till ett visst dataset, så finns det tusentals sådana dataset.

En undersökning av delning och återanvändning av data i den långa svansen av naturvetenskaplig och teknologisk forskning (Wallis, Rolando, Borgman 2013) visade att forskarna var beredda att dela sina data – om de fick erkännande för det och behöll föredradsrätt att publicera resultaten – men att få blev ombedda att göra det. Delning av data, konstaterar artikelförfattarna, drivs av efterfrågan snarare än av tillgång. Eftersom data inte dokumenteras, förses med metadata och deponeras i repositorer är de svåra att hitta och återanvända och därmed efterfrågas de sällan. Och eftersom de så sällan efterfrågas är det svårt att motivera att så mycket ansträngningar läggs på att tillgängliggöra och bevara dem. Samma typ av problem kan förklara varför det ofta saknas lämpliga repositorer för många forskningsområden i den långa svansen: för lite data delas för att berättiga

⁵ Från Chris Andersons artikel "The Long Tail" i magasinet Wired, 2004. I artikeln beskrivs hur smala nisch-produkter som tidigare sålt dåligt, genom möjligheten till exponering på internet kan öka sin andel på marknaden. Även om ingen av dessa produkter i den långa svansen säljs i så många exemplar så kan de tillsammans ta en stor del av marknaden.

att ett repositorium byggs upp och data efterfrågas för lite för att det ska vara motiverat att deponera dem i ett repositorium. I stället visade det sig att data delades i personliga kontakter med betrodda kollegor och med förväntan om att få något i gengäld. På så sätt behöll forskaren också kontroll över att delade data inte missbrukades (Wallis, Rolando, Borgman 2013).

2.1.2 Data i relation till tid

Data kan ha ett direkt och långvarigt värde, få värde med tiden eller ha ett övergående värde (Borgman 2012). Eftersom nya sätt att använda öppna data kan ge insikter som i sin tur kan resultera i ännu fler sätt att använda öppna data, förespråkar Anneke Zuiderwijk med flera en syn på öppna data som en pågående process och inte bara en produkt (Janssen, Charalabidis & Zuiderwijk 2012; Zuiderwijk et al. 2012). ”Rarely can a magic moment be established when things become data”, skriver Christine Borgman (2015, s. 62). Inspirerade av detta citat och av Borgmans tankar om att frågan ”När är data?” kanske är mer relevant att ställa än frågan ”Vad är data?”, har Jutta Haider och Sara Kjellberg (2016) undersökt föreställningar om data och tidsmässiga aspekter av hur dessa föreställningar uppstår. I fokus för studien, som omfattar både intervjuer och dokumentanalys, står arbetet med infrastrukturer för forskningsdatahantering i uppbyggnadsfasen av de storskaliga forskningsanläggningarna ESS och MAX IV i Lund.

I intervjupersonernas beskrivningar är data allt annat än statiska – snarare liknas de vid filmprojektioner och målet är att minimera tiden från utförandet av ett experiment till dess att resultatet visualiseras som data för forskarna. Idealet skulle vara att nästan helt eliminera tidsskillnaden så att datorer och mjukvara genom snabba och nästan osynliga översättningar blir ”något att se genom snarare än att se med” (Haider & Kjellberg 2016, s. 154, egen översättning). Forskarna skulle då kunna justera sina experiment som en direkt respons på de data som visas. Rörligheten är en egenskap som delas med många nya former av forskningsdata som inte existerar i en fixerad form utan endast som strömmar av observationer från exempelvis teleskop, uppkopplade sensorer, sociala nätverk och övervakningskameror: ”Any such dataset that might be shared is at best a snapshot in time” (Borgman 2012, s.1073).

När tolv forskare vid universiteten i Lund och Växjö i en intervjuundersökning (Lassi, Johnsson & Golub 2016) ombads beskriva olika stadier som deras data hade gått igenom under ett forskningsprojekt identifierade de minst två och oftast tre olika stadier: Det första stadiet omfattade rådata i form av insamlat eller erhållet material. I det andra stadiet hade materialet bearbetats, till exempel kvalitetskontrollerats, tvättats eller genomgått en första analys. Det tredje stadiet infaller vid en andra, djupare analys och när forskarna börjar ställa sin forskningsfrågor till materialet. Att dela data i första stadiet sågs som enklast och något som tio av de tolv intervjuade forskarna kunde tänka sig att göra, antingen öppet med alla eller med nära medarbetare eller andra inom fältet. De var mer tveksamma till att dela data i de två senare stadierna, och särskilt med andra än kollegor inom fältet, eftersom det finns en rädsla att data i dessa stadier kanske inte kan tolkas och analyseras på ett tillbörligt sätt (Lassi, Johnsson & Golub 2016).

Haiders och Kjellbergs intervjuer med personal på forskningsanläggningarna ESS och Max IV visade att den betydelse data förväntades ha i framtiden spelade roll för de beslut som fattades, men att det var mindre klart hur långt bort denna framtid sträckte sig:

”The accounts we were given shifted between vague hopes for a time ahead when technology will drive knowledge production and old data will be useful in ways impossible to fathom today, and more cautious, down-to-earth descriptions of technical issues, researchers, the requirements of different disciplines, and actual samples, and issues such as backing-up, file-transfers, metadata, and processing for which the future is just around the corner.

Haider & Kjellberg 2016, s. 160

I inledningen (se avsnitt [1](#)) såg vi hur nya astronomiska upptäckter möjliggjorts genom användning av tjugo år gamla arkiverade data. I artikeln (Meyer 2018) beskrivs också hur varje ny generation av observatorier är minst tio gånger känsligare än föregående generation och att det kommer att ta år för forskarna att analysera allt insamlat material. För att bevara så stora mängder data krävs inte bara mycket arbete för att förse dem med metadata och ett stort lagringsutrymme – filformat måste också vara så beständiga som möjligt eller så måste instrument och mjukvara som gör det möjligt att tolka dem bevaras.

2.1.3 Kvalitativa och kvantitativa data

Vilket syfte en forskare har haft med att samla in data och vilka metoder som använts kan påverka synen på vad som är data och vad som går att dela. Aktiviteterna runt data inom ett forskningsområde är tätt sammanbundna med de forskningsmetoder som används och det råder en dikotomi mellan kvantitativa och kvalitativa metoder (Borgman 2015). I studier som inbegriper mycket stora datamängder används ofta kvantitativa metoder och datoriserade beräkningsmodeller och de data som genereras kan vanligen anonymiseras. Detta i motsats till kvalitativa studier som kräver en noggrann analys och som kan ge djupgående beskrivningar av ett fenomen, men där det kan vara omöjligt att anonymisera – och därmed ibland även att dela – insamlade data (ibid., s. 127).

It seems likely, therefore, that the nature of disciplines, e.g. research methodologies, research problems, processes and practices within and across disciplines/sub-disciplines, shape the nature of the data created, ways in which data are manipulated and stored, the possibility of data sharing and reuse, and what constitutes effective data management practices.

Tam, Fry och Proberts, 2014, s. 723

Där det alltså i kvantitativa studier finns instrument och datorprogram som sköter datainsamlingen och genomför analysen, utgör den kvalitativa forskaren sitt eget verktyg ”för att samla in data från fältet och för att analysera det empiriska materialet” (Svensson 2015, s. 210). Detta citat väcker frågor om hur data definieras vid användning av kvalitativa metoder, och då särskilt intervjuer. I ”Handbok i kvalitativa metoder” (Svensson och Ahrne 2015) görs alltså åtskillnad mellan data och empiriskt material: det som ”produceras i själva fältarbetet” (som intervjuer och observationer) ses som data, medan det avgränsade material som sedan analyseras (exempelvis intervjuutskriften) utgör det empiriska materialet: ”Det empiriska materialet bygger på men är inte nödvändigtvis detsamma som projektets data. Detta blir kanske tydligast just i intervju- och observationsstudier, medan gränsen mellan data och empiriskt material blir mer flytande i text- och bildstudier” (Svensson och Ahrne 2015, s. 23). Ett alternativt synsätt, som vi såg i föregående avsnitt ([2.1.2](#)), är att betrakta både inspelningar och transkriptioner av intervjuer som data i olika stadier (Lassi, Johnsson & Golub 2016). Men Mats Alvesson (2011) ifrågasätter om intervjumaterial överhuvudtaget kan ses som data och hur pålitligt materialet är att bygga vidare på. ”Bara för att individer sagt vissa saker i en intervju finns inget självklart skäl till att behandla dessa som data, vilka ger starka indikationer på något som

utmärker den sociala verkligheten utanför intervjustituationen” (Alvesson 2011, s. 68). Som ett alternativ till den tekniskt inriktade synen på intervjuer där procedurerna runt insamling och bearbetning av data betraktas som huvudnumret, anser han att intervjuer måste göras till föremål för kritisk tolkning (Alvesson 2011). Skillnaderna i dessa tre olika sätt att betrakta intervjumaterial kan troligen ha stor betydelse för om, när och vilka data som delas i ett projekt. Med Alvessons synsätt – där intervjuer kanske inte ens kan ses som data för den forskare som har genomfört dem – väcks frågor om vilka möjligheter utomstående forskare har att tolka dem? Och i förlängningen även frågor om huruvida det alls är etiskt – eller ens meningsfullt – att tillgängliggöra intervjuer som öppna data?

I en undersökning av hur data både formar och formas av disciplinära kulturer jämförde Winnie Tam (2016) två underdiscipliner inom geografi: fysisk geografi och befolkningsgeografi. I de två underdisciplinerna används olika forskningsmetoder, forskare samarbetar olika ofta, i olika stora konstellationer och av olika anledningar, data delas olika ofta, av skilda anledningar och presenteras på olika sätt i publikationer. Trots att undersökningen enbart berör forskningsområdet geografi är den intressant även i ett större perspektiv eftersom många av de skillnader som Tam kartlade återfinns även mellan och inom andra discipliner. Inte överraskande visade studien att forskningsmetod och beskaffenhet hos de data som genereras och används hänger starkt ihop. Winnie Tam jämförde även skillnader i karaktär hos de data som användes, vilka former för samarbete som var vanligast, motiv och hinder för att dela data och vilka erfarenheter forskare hade av data-delning inom de respektive underdisciplinerna. För överskådlighetens skull har jag sammanställt hennes fynd i en tabell på nästa sida (se Tabell 1).

Tabell 1. Skillnader mellan två underdiscipliner inom geografi som kan påverka datadelning. Sammanställda utifrån Winnie Tams studier av hur data både formar och formas av disciplinära kulturer.

	Fysisk geografi	Befolkningsgeografi
Exempel på forsknings-ämnen	· Biogeografi, ekologiska processer. · Klimatologi, Hydrologi. · Glaciologi.	· Social och kulturell geografi. · Politisk geografi. · Utvecklings- och urban geografi.
Forsknings-metoder	<i>Enbart kvantitativa:</i> · Fältobservationer och mätningar. · Numerisk modellering. · Fjärranalys.⁶	<i>Både kvalitativa och kvantitativa:</i> · Enkätundersökningar. · Historiska källor. · Observationer och intervjuer.
Karaktär på data	· Upprepade mätningar från fältet. Numerisk modellering av inputdata, stora i volym. Fjärranalysbilder. · Innehåller sällan känslig information. · Ofta mer objektiva. · Ofta möjliga att reproducera. · Kvantitativa data, strukturerade, lättare att dela upp i olika variabler.	· Dokument från andrahandskällor. Enkät-svar. Inspelningar eller fragment av anteckningar med egna reflektioner. · Innehåller ofta känslig information. · Ofta mer subjektiva. · Ofta omöjliga att reproducera. · Kvalitativa data, ostrukturerade, måste användas i sin helhet.
Samarbets-former	· Samarbetar hela tiden. · Oftast nationella samarbeten inom samma och andra forskningsområden. · I snitt sex författare/publikation.	· Samarbetar – men inte hela tiden. · Oftast internationella och nationella samarbeten inom samma forskningsfält. · I snitt två författare/publikation.
Delning av forsknings-data	· Stor erfarenhet av att dela data utanför forskningsgruppen. · Delar sällan rådata, utan istället forskningsresultat eller delar av data.	· Liten eller ingen erfarenhet av att dela data. · Vanligare att rådata, som intervju-transkriptioner, delas.
Motiv för att dela data	· Bidra till ökad kunskap. · Ökad transparens, möjlighet att åter-skapa data och trovärdighet.	· Bidra till ökad kunskap. · Effektiv resursanvändning, undvika repetition av arbete. Nå en större publik.
Hinder för att dela data	· Konkurrens. · Att det är så tidskrävande. · Användbarheten hos delad data.	· Konkurrens. · Att det är så tidskrävande. · Användbarheten hos delad data. · Etiska och upphovsrättsliga frågor.

Källa: Winnie Tam (2016)

⁶ Fjärranalys är insamling av information om föremål eller fenomen utan fysisk kontakt med föremålet eller fenomenet. Instrumenten kan vara kameror för synligt ljus eller instrument för registrering av andra våglängder. Källa: NE.

2.2 Delning av forskningsdata

Delning av forskningsdata har definierats som "the release of research data for use by others" (Borgman 2012, s. 1060). Detta kan ske med olika grader av formalitet och tillgänglighet. Angus Whyte och Graham Pryor (2011) identifierar sex steg av öppenhet eller olika sätt att dela data: 1) *Private management* där data delas inom forskargruppen, 2) *Collaborative sharing* mellan medlemmar i ett konsortium (forskningsprojekt/ program) 3) *Peer exchange* där data delas i informella nätverk, 4) *Transparent governance* där externa parter får tillgång till data för att kunna granska och utvärdera den, 5) *Community sharing* där tillgång begränsas till medlemmar i forskarsamfundet och 6) *Public sharing* där allmänheten får tillgång till data. En genomgång av flera olika studier (Kratz & Strasser 2015) visade att en majoritet av de tillfrågade forskarna anser att datadelning är en önskvärd norm inom akademien och att många själva har hindrats i sin forskning av bristande tillgång till data. Trots det hade inte ens hälften av de forskare som var positiva till datadelning själva tillgängliggjort sina data, enligt en av de granskade studierna (Kratz & Strasser 2015).

Enligt Christine Borgman (2012) finns det fyra motiv till att dela data: 1) Att kunna reproducera eller verifiera forskning; 2) att göra forskningsresultat tillgängliga för alla; 3) att göra det möjligt för andra att ställa nya frågor till existerande data och 4) att främja forskning. En annan fördel är att onödiga datainsamling undviks när redan insamlad data kan återanvändas, något som sparar både tid och pengar (Callaghan et al. 2012).

For this reason, research funders are keen to encourage data sharing. The tension on the researchers' side is that there is (currently) no universally accepted mechanism for data creators to obtain academic credit for their dataset creation efforts. Consequently, they often prefer to hold the data until they have extracted all the possible publication value they can. Though completely understandable, this behaviour comes at a cost for the wider scientific community.

Callaghan et al. 2012, s.108

Ett problem är alltså att de som förväntas investera tid och pengar i att göra data tillgängliga inte nödvändigtvis är de som har mest att vinna på att data delas. Till skillnad från vetenskapliga publikationer, som är dokument skapade för att kunna spridas och stå för sig själva, måste nämligen data sättas i sin kontext (Borgman 2015; Mauthner & Parry 2013): Om data ska vara möjliga att återvända måste de organiseras, förses med metadata och dokumenteras så att de går att hitta och förstå – det räcker inte att bara tillgängliggöra dem i ett arkiv eller repositorium. Att skapa denna kontext tar tid och flera undersökningar av forskares attityder till datadelning har visat att just tidsbrist är den enskilt viktigaste förklaringen till varför data inte delas (Kratz & Strasser 2015; Tenopir et al. 2011). Andra orsaker är brist på pengar, att forskarna inte har eller tror sig ha rättigheter att tillgängliggöra data offentligt, att de saknar standarder eller ett ställe att förvara data på, att finansieringen inte kräver att data delas eller att forskarna anser att deras data inte bör vara tillgängliga av exempelvis etiska skäl (Tenopir et al. 2011). Det finns också en oro över att andra forskare ska missbruka öppenheten, vilket var en av förklaringarna till att även forskare som själva har använt sig av andras data kan ställa sig tveksamma till att fritt tillgängliggöra sina egna data (Hickson, Poulton, Connor, Richardson & Wolski 2016; Tenopir et al. 2011). Som tidigare nämnts (se avsnitt [2.1.1](#) och avsnitt [2.1.3](#)) kan praktikerna runt datadelning också variera kraftigt mellan – och även inom – olika discipliner.

2.2.1 Policyer och stöd för forskningsdatahantering

Trots att allt fler finansiärer och lärosäten ställer krav på att data ska tillgängliggöras öppet är detta långt ifrån verklighet för alla forskare inom alla discipliner (Elsevier & CWTS 2017). I den mån data alls delas är det fortfarande vanligt att det sker informellt genom personliga kontakter, och utvecklingen runt infrastruktur, utbildning, organisation och finansiering för och av datadelning har inte hunnit med (ibid.).

Jens Klump (2017) beskriver hur kulturen i forskarvärlden, där medlemmarna i olika grupper ger varandra gåvor med en förväntan om att detta ska återgäldas i framtiden, kan vara ett hinder för öppen datadelning. Gåvorna kan bestå av exempelvis en inbjudan till att tala på en konferens eller tillgång till instrument – eller data. När det gäller tidskriftsartiklar byts information mot erkännande. ”Att lägga ut data på internet till anonyma användare utan att kunna vänta sig en gåva i gengäld är inte en drivkraft i denna modell av vetenskapliga kommunikation eftersom det bryter mot principen om ömsesidighet som är fundamental för gåvokulturen” (Klump 2017, s. 3, egen översättning). Men enligt Veerle Van den Eynden och Libby Bishop (2014) finns det flera viktiga motiv för forskare att dela sina data och inte alla innefattar en direkt eller indirekt belöning. Dessa motiv är:

(1) when data sharing is an essential part of the research process; (2) direct career benefits derived from sharing through greater visibility of one’s work, reciprocal data exchanges, and the reassurance of having one’s data recognised as valuable by others; (3) the norms that researchers are exposed to within their research circle or discipline; and (4) a framework of funder and publisher expectations, policies, infrastructure and data services as external drivers.

Van den Eynden & Bishop 2014, s. 4

I sin undersökning intervjuade Van den Eynden och Bishop 22 forskare i fem forskningsgrupper i Danmark, Finland, Tyskland, Nederländerna och Storbritannien, alla från olika discipliner och med olika kulturer runt delning av data. De fann att även om normer i de skilda disciplinerna var viktiga så fanns det stora variationer också inom disciplinerna – och mönster som sträckte sig över disciplingränserna. Forskarna hade också olika skäl att dela data under olika skeden av sin karriär. Att normer runt datadelning har ett starkt inflytande tyder, enligt författarna, på att tidig utbildning om datadelning, till exempel som en del av metodikutbildningen, kan ha en stor roll. De argumenterar också för vikten av policyer för öppna data – ”så länge de inte förhindrar informell delning av data och tar hänsyn till variationer mellan disciplinerna (Van den Eynden & Bishop 2014, s. 5, egen översättning):

Policies operate through at least two channels. They provide a collective voice and help to clarify and change the norms of the research community. Policies also alleviate a mismatch of incentives: they create a positive motivation to share that benefits science in general, even in those cases where the direct benefits to individual researchers are weaker.

Van den Eynden & Bishop 2014, s. 5

Natasha Mauthner och Odette Parry (2013, s. 48) ifrågasätter bilden av att policyer för öppna forskningsdata är passiva, neutrala och per automatik främjar forskning. Motstånd mot att följa dessa policyer förenklas då till en rent teknisk fråga om att förbättra infrastrukturen för delning av forskningsdata. I stället pekar Mauthner och Parry på det problematiska i att göra forskningsdata till allmänna resurser som är oberoende av de mänskliga, sociala och materiella förhållanden under vilka de skapats. Även i artikeln ”If We

Share Data, Will Anyone Use Them?” (Wallis, Rolando & Borgman 2013) väcks frågan om huruvida policyer och tvång verkligen är rätt sätt att öka tillgången till öppna forskningsdata. Att som forskare investera så mycket tid i att dokumentera sina data bara för att en okänd användare eventuellt skulle kunna få glädje av dem någon gång i framtiden är inte rimligt, skriver författarna. Och att kurerat alla dessa data så att de förblir tillgängliga på lång sikt blir också orimligt dyrt för exempelvis universitet eller repositorer.

Thus, the better question to ask is which data are worth the investment for reuse? This is yet another chicken-and-egg problem, unfortunately. Data for which demand exists should be kept and curated. Demand arises only if data are discoverable, accessible, and usable. How can demand be built until a critical mass of useful data are available?

Wallis, Rolando & Borgman 2013, s. 15

Studier om forskares attityder till, och arbete med, delning av data är av stor betydelse, inte minst för att universiteten ska kunna erbjuda rätt stöd och bygga upp infrastrukturer i olika faser av forskningsdatahantering. Här har universitetsbiblioteken en viktig roll. I Sverige har sådana studier gjorts vid bland annat universiteten i Lund och Malmö, för att ta två närliggande exempel. Jag kommer här att ta upp tre av dessa eftersom de berör frågor som har särskild relevans för uppsatsens frågeställningar.

I Lund har intervjuer genomförts både med representanter från de olika fakulteternas ledning och bibliotek (Johnsson & Lassi 2016) och senare – i samarbete med Linnéuniversitetet i Växjö – med forskare (Lassi, Johnsson & Golub 2016, se även avsnitt [2.1.2](#)). Den första undersökningen visade på stora skillnader mellan forskningsområden i arbetet med att tillgängliggöra forskningsdata. Det framkom också att det är vanligt att forskare tror att det är de själva som äger data, inte lärosätet eller myndigheten, och att det ”kan bli en pedagogisk utmaning för lärosäten att förklara för forskare att forskningsdata är att betrakta som en allmän handling vilken kan begäras ut av allmänheten och att forskningsdata tillfaller lärosätet” (Johnsson & Lassi 2016, s. 13). I rapporten ges en rad förslag till hur arbetet med hantering av forskningsdata kan utvecklas framöver. Bland annat rekommenderas en översyn av meriteringssystemet så att tillgängliggörande och andras citeringar av forskningsdata inkluderas. Intervjuerna med forskare (sju i Lund och fem i Växjö) bekräftade bilden av att det finns stora variationer mellan forskningsområden och att forskare inom områden som kräver stor noggrannhet vid datainsamlingen inte oväntat hade mer utvecklade och grundliga rutiner för forskningsdatahantering. De intervjuade forskarna var överlag positiva till att dela data och utbyte av data var vanligt förekommande, särskilt inom olika samarbeten. Det fanns också en medvetenhet bland de intervjuade forskarna om vilka data som kan vara svåra att dela av etiska skäl (Lassi, Johnsson & Golub 2016). Med tanke på att urvalet i den senare undersökningen gjordes bland annat genom e-postefterlysningar och personliga kontakter är det dock svårt att uttala sig om representativiteten i svaren: det är möjligt att forskare som av olika anledningar redan är engagerade i arbetet med delning av data har varit mer benägna att ställa upp än forskare som inte är det.

I Malmö har flera olika studier om datahantering och forskarstöd gjorts och jag vill särskilt nämna en enkätundersökning där universitetets forskare fick svara på frågor om sitt arbete med forskningsdata (Fransson, du Toit, Kjellberg & Tapia Lagunas 2016; se även Fransson et al. 2017). Enkäten visade att en majoritet av de svarande använder andras data och även delar sina egna. Få forskare har dock hittills tillgängliggjort sina data öppet

i exempelvis dataarkiv. Egen delning sker i första hand genom informella kontakter och de vanligaste svaren på frågan om vad som skulle få forskarna att börja eller i större utsträckning göra data öppet tillgängliga är ”efterfrågan av forskare i samma ämne” samt ”avsatt tid”. Nästan en av tre svarade att det skulle vara en drivkraft om tillgängliggörande av data sågs som en merit. Även krav från finansiärer skulle ha betydelse för cirka en tredjedel, medan en policy på Malmö högskola bara skulle motivera 14 procent av respondenterna. Svaren visade också hur forskningsetiska och juridiska aspekter kan krocka och att det finns en oro för att datahanteringen kommer att ta av forskningstiden. Forskarna efterlyste mer stöd omkring arbetet med forskningsdata i alla faser (Fransson, du Toit, Kjellberg & Tapia Lagunas 2016). Även här är det dock svårt att uttala sig om representativiteten eftersom svarsfrekvensen var relativt låg (27 procent) och det inte går att veta vilka grupper som valde respektive valde bort att svara på enkäten.

2.2.2 Etik och integritet

Kraven på öppna data kan innebära komplicerade etiska överväganden. I en studie rörande genetikforskning (Zarate et al. 2016) beskrivs till exempel hur personuppgifter kan vara svåra eller omöjliga att redigera när de väl har delats på internet. Genetikforskningen är speciell eftersom all data är unikt personlig och nya metoder har möjliggjort återidentifiering av data som tidigare ansetts vara avidentifierade.

[...] identities may be revealed by a process called data linkage, in which online data sets are matched to each other. Reidentification (re-ID), the process of associating an individual's name with data that were considered deidentified, poses risks such as insurance or employment discrimination, social stigma, and breach of the promises often made in informed-consent documents.

Zarate et al. 2016, s. 36

Och det är inte bara genetikforskare som behöver fundera över dessa risker, enligt studien. Även andra typer av hälsorelaterade data och data från social forskning är känsliga för datalänkning och återidentifiering. Nya modeller för informerat samtycke behövs därför, enligt författarna, för all typ av forskning som rör människor. Om data ska delas räcker det inte att informanterna samtycker till att delta enbart i den aktuella forskningsstudien. Och om det konfirmerade samtycket i stället formuleras så öppet att deltagarna måste godkänna även all framtida användande av deras uppgifter – av okända forskare och på sätt som kanske inte ens går att förutse i dag – riskerar detta att undergräva tilliten till vetenskapen och viljan att delta i forskningsprojekt. Ett alternativ till ett sådant öppet samtycke skulle kunna vara dynamiskt samtycke, vilket innebär att deltagarna kan ha upprepade kontakter med forskare där de kan tillåta att vissa av deras uppgifter används för specificerade analyser (Zarate et al. 2016).

Med anledning av EU:s dataskyddsförordning (GDPR) som började gälla den 25 maj 2018 tillsatte regeringen en särskild forskningsdatautredning med uppdrag att lämna förslag till kompletterande bestämmelser. Syftet var att ”möjliggöra en ändamålsenlig behandling av personuppgifter för forskningsändamål samtidigt som den enskildes fri- och rättigheter skyddas” (Forskningsdatautredningen 2017). Utredningen var ännu inte färdig när denna undersökning färdigställdes, men i ett delbetänkande identifierades behovet av att införa en särskild forskningsdatalag som ska komplettera dataskyddsförordningen. Återanvändning av personuppgifter som samlats in för forskningsändamål ska, enligt delbetänkandet, även fortsatt vara möjlig så länge det finns rättslig grund för att uppgif-

terna behandlas igen. Forskning av allmänt intresse kan räknas som en sådan rättslig grund. Normalt ska också samtycke inhämtas på nytt (Forskningsdatautredningen 2017).

2.3 Publicering av data

De senaste åren har fokus flyttats från *delning* av data till *publicering* av data (Kratz & Strasser 2014). Att publicera har enligt ordböckerna betydelsen ge ut i tryck, men grundbetydelsen är att göra något publikt tillgängligt, offentliggöra.⁷ Digitaliseringen har gjort det enkelt för vem som helst att publicera ett dataset genom att göra det tillgängligt på en webbsida på internet (Lawrence, Jones, Matthews, Pepler & Callaghan 2011). Men åsikterna om vad som avses med publicering av data varierar kraftigt.

Data publication overlaps with terms like data sharing, data release, and open data. A data publication might be a spreadsheet on a website, a set of images in an institutional archive, a stream of readings from a weather station transmitted over the internet, or a peer-reviewed article describing a dataset ...

Kratz och Strasser 2014, s. 3

Data kan till exempel publiceras av utgivaren för en (data)tidskrift, i en databas, på en forskares webbsajt eller i ett repositorium eller arkiv. Ett publicerat dataset kan antingen vara ett komplement till en konventionell forskningsartikel, vara själva ämnet för en dataartikel eller dokumenteras fristående av sin utgivare (Kratz & Strasser 2014).

I följande två avsnitt kommer jag att ta upp viktiga funktioner som olika sätt att publicera data kan ha, och särskilt då på dataartikelns och datatidskriftens betydelser.

2.3.1 Olika sätt att publicera data och publiceringens betydelse

När ett dataset publiceras på internet blir det ofta öppet tillgängligt, men det är inte säkert att det finns tillräckligt med metadata för att det ska gå att hitta och vara begripligt för en utomstående eller att det finns kvar på lång sikt och i ett format som går att använda. Och även om ett dataset har tillräckligt med metadata och är tekniskt användbart för en utomstående finns det ingen information om dess vetenskapliga kvalitet; hur data publicerats och av vem, hur publicerad data dokumenterats, var dokumentationen finns i förhållande till de data som beskrivs och hur data har validerats (Callaghan et al. 2012, s. 108). Bryan Lawrence och Sarah Callaghan med flera vill därför skilja mellan *publicering* – vars begränsningar beskrivits ovan – och *Publicering* med versalt P (Callaghan et al. 2012; Lawrence et al. 2011). Att *Publicera* data definieras då som: ”Att göra data så permanent tillgängliga som möjligt på internet” (Lawrence et al. 2011, s. 7, egen översättning⁸). *Publicerad* data ska vidare, enligt författarna, ha genomgått en process där den förses med lättbegriplig information om tillförlitlighet, innehåll och format. Publicering med versalt P ger alltså ett dataset ökat värde för framtida användare. Detta kan ske genom att datasetets vetenskapliga kvalitet och betydelse bedöms i en peer review-process, att det förses med tillräckliga metadata och att dess tekniska kvalitet säkras (Callaghan et al. 2012). Sådan Publicering erbjuder också dataproducenterna en möjlighet att få akademiskt erkännande för sitt arbete (ibid.).

⁷ [A]v lat. *publicare* 'offentliggöra', till *publicus* 'allmän'. Källa: Nationalencyklopedin.

⁸ ”To make data as permanently available as possible on the Internet”

Utgångspunkten är de fem funktioner i publiceringsprocessen som identifierats av Herbert Van de Sompel med flera (Van de Sompel et al. 2004 se Lawrence et al. 2011, s. 8): *Registrering* (så att en forskare kan hävda företräde till en upptäckt), *certifiering* (för att fastslå giltigheten i den vetenskapliga upptäckten), *synliggörande* (så att aktörer inom akademien kan hålla sig uppdaterade om nya upptäckter), *arkivering* (för att bevara vetenskapliga dokument över tid), samt *belöning* (så att prestationer i kommunikationssystemet belönas utifrån olika mätmetoder). När det gäller publicering av data lägger Lawrence med flera till en sjätte funktion, nämligen *definition* (av vad det är som har blivit publicerat). Ingen av dessa funktioner säger dock något om kvalitetskontroll – en funktion som inom akademien uppfylls av peer review-processen. Författarna argumenterar därför för att samma acceptans för peer review-processen som kvalitetssäkring nu behövs vad gäller datapublikationer. Men vad innebär peer review av data? Enligt artikeln ”Data Citation and Peer Review” (Parsons, Duerr och Minster 2010) saknas entydiga definitioner. Är det själva datans exakthet och giltighet som har granskats eller är det dokumentationen som har synats för att garantera att kontext och eventuella osäkerheter har dokumenterats? Författarna anser att det kan vara problematiskt att citera en artikel som beskriver ett dataset i stället för att citera själva datasetet. Särskilt gäller detta i fall när datasetet fortsätter att utvecklas efter det att dataartikeln publicerats. I stället föreslår de ett sätt att citera data som liknar det forskare använder när de citerar monografier, med författare, redaktör, publiceringsdatum och version/upplaga, utgivare, datasetets titel och en tydlig hänvisning till var det finns. Som författare vill de då se personerna som har gjort tankemässiga ansträngningar för att samla in och göra i ordning datasetet, oavsett om de är datainsamlare, teamledare, algoritm-utvecklare eller har någon annan roll (Parsons, Duerr & Minster 2010).

2.3.2 Datatidskrifter och dataartiklar

Traditionellt – och något förenklat – har vetenskapligt arbete bedömts utifrån de slutsatser som har dragits av insamlade data och inte utifrån kvaliteten på dessa data. Ansträngningarna att producera data har sällan värderats högt trots att uppgiften ofta är tidskrävande. Datapublikationer i olika former fyller här en funktion eftersom de bidrar till att synliggöra det arbete som forskare lägger ner på att skapa data (Lawrence et al. 2011). Exakt vad som kvalificerar en datapublikation är ännu inte helt klargjort. Viss enighet råder dock om att publicerade data måste vara så väl dokumenterade att åtminstone forskare inom samma område ska kunna återskapa eller återanvända dem, de ska kunna citeras på samma sätt som en tidskriftsartikel och de ska finnas permanent tillgängliga även om de inte behöver vara helt fria att använda⁹ (Kratz och Strasser 2014). En fjärde egenskap som diskuterats rör frågan om data även måste valideras – och i så fall hur – för att räknas som publicerad, men Kratz och Strasser nöjer sig med att räkna data som publicerad så länge den är beständigt tillgänglig, dokumenterad och möjlig att citera. Det är svårt att hitta en exakt siffra över antalet datatidskrifter, men enligt en uppskattning från 2016 (Akers) rörde det sig då om cirka tvåhundra titlar. Dessa kan delas in i renodlade (”pure”) datatidskrifter och blandade (”mixed”) tidskrifter som accepterar dataartiklar som en av flera typer av artiklar (Akers 2016). I allmänhet innehåller en dataartikel inga hypoteser eller analyser utan syftet är att beskriva ett eller flera dataset, deras kontext, kvalitet, vilka

⁹ En avgift eller krav på juridiskt avtal för att använda data kan accepteras så länge detta är tydligt reglerat och inte styrs av författarens godtycke (Kratz and Strasser 2014).

mätmetoder som använts, hur instrument kalibrerats och liknande. Eventuellt kan författaren också ge förslag på hur dessa dataset skulle kunna användas av andra i framtida forskning. Tre fördelar med dataartikeln är att den 1) är en publikation som kan sättas upp på ett forskar-cv på samma sätt som konventionella artiklar och därmed bidra till att författaren/dataproducenten får erkännande för sitt arbete, 2) fungerar som en slags förlängd metadata som beskriver data på ett strukturerat sätt och 3) sprider kännedom om dessa data till forskarsamfundet (Chavan & Penev 2011).

Data papers and joint data-publishing and -archiving policies by databases and journals in combination address a cluster of related requirements: the persistent, stable, long-term archival of data; a better standardization of data sets and improved confidence about their quality and fitness for use; reliable linkages between scholarly articles and the underlying data; and global acceptance that research data should be readily accessible and reusable.

Chavan, Penev & Hobern 2016

Till detta kan läggas att en dataartikel kan öka möjligheterna att hitta och återanvända data, även sådana som inte är kopplade till en konventionell vetenskaplig artikel. När obearbetade dataset blir en viktig del av den datadrivna forskningen är dokumentation en viktig del av forskningsdatahanteringen för att möjliggöra att dessa data kan återanvändas (Ya-Ning 2016). Dataartiklar som speglar den konventionella vetenskapliga publikationsmodellen kan då utgöra ett komplement till metadata (ibid.). Andra fördelar är att en dataartikel ofta granskas i en peer review-process vilket ger det beskrivna datasetet ökad legitimitet. Citeringsförfarandet kan också underlättas, eftersom det annars ofta är oklart hur ett dataset ska citeras. En dataartikel kan citeras i datasetets ställe och enligt samma princip som en konventionell forskningsartikel. Citeringarna kan sedan ofta spåras på samma sätt som citeringar till konventionella artiklar. Det har visat sig att tidskriftsartiklar som kompletteras med data citeras mer än artiklar utan data (Piwowar & Vision 2013). Ofta används dessutom dessa data i ny forskning vilket innebär att upphovspersonerna citeras i så kallade *data reuse papers* (ibid.).

Den formaliserade citeringsprocessen innebär också att datadelning synliggörs när bibliometrisk metod används vid forskningsutvärdering och fördelning av forskningsmedel. I Sverige förekommer detta sedan 2009 då regeringen införde prestationsbaserad omfördelning till svenska lärosäten baserad bland annat på publikations- och citeringsmått (Nelhans & Eklund 2015). Bibliometrisk indikator används då för att fördela forskningsmedel såväl mellan olika lärosäten som mellan ett lärosätes olika institutioner. ”I vissa fall utnyttjas bibliometrisk indikatorer ner på individnivå, där forskare antingen tilldelas resurser i andel arbetstid avsatt för fortsatt forskning, i specifika excellenspotter eller som bonus i form av ekonomisk ersättning för prestation” (Nelhans & Eklund 2015, s. 25). Kritiker av systemet har påtalat att det kan medföra en risk att forskare – hellre än att ägna sig åt nydanande forskning som kan ha lång genomslagstid – går in för att maximera sina rankingpoäng (Nelhans & Eklund 2015, s. 20).

Det är också viktigt att hålla i minnet att datatidskrifter är ett sätt för de vetenskapliga förlagen att hålla fast vid en lönsam affärsidé där forskarna gratis levererar och granskar artiklar och dessutom numera i många fall måste betala särskilda avgifter ¹⁰för att artik-

¹⁰ Så kallade Article Processing Charges eller APC:er, se förklaring i avsnitt [1.4](#)

larna ska göras öppet tillgängliga (Fagerström 2016). Detta är särskilt intressant med tanke på att de vetenskapliga förlagens affärsidé är starkt ifrågasatt i och med att allt fler länder vägrar fortsätta betala dyra prenumerationsavtal. Medan denna uppsats skrevs sades till exempel de svenska forskningsbibliotekens avtal med det stora vetenskapliga förlaget Elsevier upp på grund av att ”förlaget inte tillgodoser kraven på öppen tillgång” (Kungliga biblioteket 2018). Vitek Tracz och Rebecca Lawrence (2016) framför i en debattartikel flera argument för att tidskrifter överlag håller på att bli en utdaterad modell för publicering av nya forskningsrön, bland annat för att de tar in så få artiklar att det försenar tillgängliggörandet av forskning och att granskningsförfarandet inte är transparent. Strävan efter hög Journal Impact Factor innebär också att mycket forskning i mindre studier eller studier med negativa eller neutrala resultat aldrig publiceras eftersom tidskrifterna premierar artiklar som förväntas ge många citeringar.

A new way of publishing, discussing and reviewing new scientific findings is urgently needed to speed up the progression of science, and to improve the fairness of the system used to judge researchers with regards to their next grant or career move. Such a new system should also enable funders to maximise the value of their research investment.

Tracz & Lawrence 2016, s. 2

I stället föreslår författarna en plattform för öppen vetenskap som utnyttjar fördelarna hos preprint-formatet, men lägger till en transparent granskningsprocess. Detta skulle flytta fokus i utvärderingen av forskning och forskare från att mäta genomsnitt på tidskriftsnivå (Journal Impact Factor, se förklaring i avsnitt [1.4](#)) till ett mer rättvist system där kvalitet bedöms på artikelnivå, hävdar Tracz & Lawrence (2016).

3. Teoretiska perspektiv

Det krävs många slags expertkunskaper för att hantera data i olika stadier av forskningsprocessen: teoretiska, praktiska, sociala och tekniska (Borgman 2015, s. 36). Synen på kunskap och forskning som ett socialt fenomen och som något som skapas genom de praktiker som uppstår ur människors dagliga aktiviteter är utgångspunkten för denna uppsats.

Scientific facts are not found, literally dis-covered, in nature, but they are actively construed by scientists. Readings from instruments do not speak for themselves but need to be constructed into scientific facts by researchers. The processes in which this is accomplished are social by their very nature: human researchers interacting with each other. They cannot be understood as mechanically following methodological rules; if that were so, we could replace scientists by computers.

Bijker 2001, s. 24

För att undersöka studiens frågeställningar använder jag vidare ett sociotekniskt perspektiv, där varken tekniska eller sociala aspekter sätts främst, utan båda ses som ömsesidigt beroende av varandra. Jag utgår också från ett sätt att betrakta dokumentet där fokus flyttas från synen på information som abstrakt begrepp till dokumentets materiella aspekter (Francke 2008). I följande tre avsnitt kommer jag att beskriva dessa teoretiska utgångspunkter; forskares sociala praktiker, det sociotekniska perspektivet och dokumentets betydelse.

3.1 Sociala praktiker och epistemiska kulturer

”People learn and work within groups that share understanding, practices, technology, artefacts and language”, skriver Nancy Van House (2003a, s. 283) utifrån Jean Lave och Etienne Wengers syn på lärande som tätt sammankopplat med såväl sociala praktiker som materiella fenomen. I teorier om sociala praktiker betonas det ömsesidiga beroendet mellan individens agerande och omvärlden (Van House 2003a, Van House 2003b):

Practice is people’s actual, daily, embodied activity, often including skills, tacit knowledge, and presuppositions, as well as their interaction with others and with material and other resources. Specific practices are generally understood to differ to some degree across domains.

Van House 2003b, ss.41-42

En praktikorienterad ansats kan användas för att ”fokusera på relationen mellan omvärlden och individens agerande och hur detta sker i ett visst sammanhang, samt dessutom i samspel med olika verktyg” (Kjellberg 2016, s. 50). Dessa sammanhang kan skilja sig åt inom olika discipliner och forskningsområden, men också ha likheter som överskrider disciplingränserna (Van House 2003b). Fenomenet har studerats av bland andra Karin

Knorr Cetina (1999) som myntade begreppet *epistemiska kulturer* för att beskriva hur olika "arrangemang och mekanismer" inom ett givet fält "avgör *hur vi vet vad vi vet*" (Knorr Cetina, 1999, s. 1, författarens betoning, egen översättning).

I boken "Epistemic cultures" (1999) undersöker hon hur kunskap skapas genom att jämföra två olika laboratoriemiljöer inom forskningsfälten högenergifysik och molekylärbio-logi. Bland annat tittar hon på hur förtroende byggs upp mellan människor som samarbetar (ibid., ss 130-134 & ss 201-209). Graden av förtroende forskarna inom högenergifysik känner för en kollega avgörs till exempel av ett antal olika faktorer, både formella och informella. Det kan röra sig om hur långt kollegan har kommit i sin akademiska karriär, om denne är fast anställd eller inte, expert eller icke-expert, men också om vem som får saker gjorda eller mest är fulla av tomma ord. Medarbetarna i kollaborativa experiment delas in i grupper utifrån "vad som är känt om dem; vems arbete som går att bygga vidare på, vems resultat är ' trovärdiga', och vem vill man 'samarbeta med' och, alternativt, vem vill man undvika" (Knorr Cetina 1999, s.131, egen översättning). Knorr Cetina beskriver hur *confidence pathways* (ibid., s. 202) utvecklas mellan forskare som delar historia genom experiment de har deltagit i. Dessa nätverk skapar en egen form av organisation som sträcker sig utanför institutioner och lever kvar även efter experimentets slut.

Nancy A. Van House (2003b) förklarar hur begreppet epistemiska kulturer kan vara användbart inom informationsvetenskaplig forskning av flera skäl: Kunskap sammankopplas med materiella fenomen såväl som sociala praktiker, betydelsen av aktiviteter och individers dagliga arbete betonas, och mångfald och olikhet kan studeras genom att discipliner och epistemiska kulturer särskiljs.

3.2 Dataartikeln som socioteknisk ensemble

Det sociotekniska perspektivet har, som nämnts tidigare (se avsnitt [1.3](#)), sitt ursprung i teknik- och vetenskapsstudier (Van House 2003b; Park 2008). Ett sociotekniskt system skapas och upprätthålls i samspel mellan människor, teknik och praktiker, där tekniken sätter ramarna för vad människor kan göra, samtidigt som människors behov och beteende driver fram en ständig utveckling av tekniken. Den påverkan som modern informationsteknologi har på forskning och vice versa beskrivs på liknande sätt av Christine Borgman som en push and pull-effekt:

The production of more digital content is pushing the development of scholarly information infrastructure technologies to manage it, and the availability of more digital content, tools and services is pulling more scholars toward using them.

Borgman, 2007 s. 31

Social Construction of Technology (SCOT) utgår från relevanta sociala grupper och dessa grupper syns på tekniska artefakter, där samma artefakt kan få helt olika betydelse för de olika grupperna. Wiebe E. Bijker använder begreppet *interpretative flexibility* för att visa hur "neither an artifact's identity, nor its technical 'success' or 'failure', are intrinsic properties of the artifact but subject to social variables" (Bijker 2001, s. 26). Stabiliteten hos ett sociotekniskt system är alltså inget som kan tas för givet utan är ett resultat av ständigt pågående förhandlingar där varken den sociala aktören eller den tekniska artefakten är intressanta att studera i sig. I stället riktas fokus mot de nätverk av ömsesidigt stabiliserande aktiviteter och objekt som uppstår och som bland annat har kallats för

sociotekniska ensembler (Bijker 1995, Levy 2003). Alla stabila ensembler, hävdar Bijker, sammanbinds lika mycket av tekniska som sociala faktorer. Genom att utgå från ett sociotekniskt perspektiv har elektroniska open access-tidskrifters utformning tidigare studerats av Helena Francke (2008) och forskarbloggar av Sara Kjellberg (2010). Helena Francke (2008) argumenterar, utifrån Bijker, för hur dokumentet och i synnerhet då den vetenskapliga artikeln, kan ses som en socioteknisk ensemble eftersom varken dess utformning eller de praktiker som uppstår runt en vetenskaplig artikel kan förklaras av enskilda tekniska eller sociala faktorer.

The stable artefact, in this view, does not come about as a result of necessity or because it is the "best solution". Rather, it is the outcome of a process of negotiation, where the interests of different user groups and technological solutions are engaged in deliberations on how the artefact should be interpreted, designed, and used.

Francke 2008, s. 106

På samma sätt ser jag dataartikeln som en socioteknisk ensemble, som påverkas både av olika grupper av användare, till exempel de forskare som skriver dataartiklar och de som citerar dem, och av tekniska förutsättningar, som olika sätt att dela och publicera data. Som litteraturgenomgången i kapitel 2 visade finns det många instabila processer runt delning av data (se till exempel avsnitt [2.2.1](#), [2.2.2](#) samt [2.3.1](#)). Genom att betrakta dataartikeln som en socioteknisk ensemble kan jag rikta fokus mot de sociala och tekniska faktorer som påverkar hur den utformas och vilka praktiker den ger upphov till. Ur en vidare synvinkel skulle även datatidskriften kunna ses som sådan ensemble, där användarna också inkluderar exempelvis redaktörer på vetenskapliga förlag och prenumeranter medan de tekniska förutsättningarna kan gälla olika sätt att publicera och finansiera tidskrifterna. Detta ligger dock utanför ramarna för denna uppsats.

Möjligheten till social analys av teknik erbjuder dessutom ett konstruktivistiskt perspektiv som, enligt Bijker (2001), utgör en rationell grund för teknikpolitik:

Demonstrating the interpretative flexibility of an artifact makes clear that the stabilization of an artifact is a social process, and hence subject to choices, interests, value judgements – in short, to politics. Without recognizing the interpretative flexibility of technology, one is bound to accept a technologically determinist view.

Bijker 2001, s. 27

Detta är intressant, särskilt med tanke på de policyer för öppna data som för närvarande håller på att utvecklas och implementeras på olika nivåer, och hur sådana policyer kan påverka forskare i deras arbete.

Påpekas bör slutligen att även om det sociala (samhället) och teknologin, enligt sociotekniska teorier, ses som mycket tätt sammanflätade måste de inte alltid behandlas som en enhet: ".../ analytical convenience may demand that, in accounts of technology, the technical and the social be treated as separate concepts" (Kling, McKim & King 2003, s. 55). Denna möjlighet att först se på de båda faktorerna var för sig för att sedan koppla samman dem igen, kommer jag att använda mig av i analysen.

3.3 Dokumentets roll

Att sprida och få tillgång till vetenskaplig information sägs ofta vara en av huvudfunktionerna i den vetenskapliga kommunikationsprocess som har utvecklats genom århundraden och där systemet med vetenskapliga tidskrifter fått acceptans som ”det snabbaste, bekvämaste och mest tillförlitliga sättet att sprida och få tillgång till vetenskapliga forskningsfynd” (De Silva & Vance 2017, s. 20, egen översättning). I boken ”Deflating Information” ifrågasätter Bernd Frohmann (2004) synen på den vetenskapliga artikeln som ren bärare av information eller epistemisk kunskap – i alla fall den typ av information som behövs för att producera ny kunskap. En sådan epistemisk diskurs skapar nämligen, enligt Frohmann, en paradox i det att så många vetenskapliga artiklar aldrig citeras och att forskare verkar föredra att utbyta information genom informella kanaler. Därmed inte sagt att vetenskapliga artiklar inte överför någon information alls, eftersom de kan ge en översiktlig bild av ett forskningsfält, och bekräfta kunskapsläget för den som vill placera ett vetenskapligt arbete i relation till ett annat.

En viktig roll för den vetenskapliga artikeln är också att verka stabiliserande, eftersom den medverkar till att placera in författaren i en social hierarki och blir en del av det akademiska belöningsystemet (Frohmann 2004). Även Sabina Leonelli betonar de vetenskapliga tidskrifternas stabiliserande roll inom akademien, i det att de avgränsar vetenskapliga gemenskaper och delområden, och kan vara ett skyltfönster för centrala insikter (Leonelli 2016, s. 393). Den vetenskapliga artikeln verkar dessutom stabiliserande i sig genom att datainsamlings- och analysprocesserna passas in i en sammanhängande berättelse. Målet är att övertyga läsarna om de slutsatser som dragits. Aspekter av forskningen som inte understödjer argumenten utelämnas – det kan exempelvis röra sig om misslyckade experiment eller modeller, data som visat sig vara irrelevanta eller som är neutrala i förhållande till uppsatta hypoteser (Leonelli 2016).

I stället för att se den vetenskapliga artikeln som bärare av abstrakt, immateriell kunskap argumenterar Frohmann för ett synsätt där fokus, på ett liknande sätt som i teknik- och vetenskapsstudier, läggs på aktiviteterna och de materiella aspekterna av dokumentet (Van House 2003b). Dokumentperspektivets fokus på materialitet pekar på svårigheterna med att betrakta information som kontextlös och betonar vikten av ”den tid och plats där dokumentet skapas, hanteras och studeras” (Francke 2008, s. 97, egen översättning).

In discourses of practice, activities are varied, skilful, craftlike, and local; they engage a wide variety of material, social, and discursive objects, with the aim of producing more-or-less stable assemblages of things, people, and documents, which can be extended beyond the laboratory as resources for other research contexts and transscientific fields.

Frohmann 2004, s. 102

Praktikdiskurser passar därmed bra för att undersöka dataartikeln, som ju lägger fokus på resultat och metoder snarare än teorier och argumentation. Ett av dataartikelns huvudsakliga syften är dessutom just att vara en resurs i andra forskningskontexter.

3.4 Sammanfattning

För att undersöka studiens frågeställningar om dataartikeln och dess betydelse i forskares arbete utgår jag alltså från att forskning och vetenskaplig kommunikation är en process

som utformas i samspel mellan människor, teknik och artefakter och att denna process ser olika ut i olika grupper. Därmed anknyter uppsatsen till tidigare studier inom biblioteks- och informationsvetenskap där skilda aspekter av vetenskaplig kommunikation har studerats som lokalt situerade i olika praktiker och ur ett sociotekniskt perspektiv. Detta teoretiska ramverk blir min utgångspunkt för att analysera intervjusvaren och besvara forskningsfrågorna.

4. Intervju som metod

Både det teoretiska ramverket och forskningsfrågorna är viktiga vid val av metod. Min utgångspunkt är att vetenskaplig kommunikation är en process som utformas i samspel mellan människor, teknik och artefakter, och vars praktiker ser olika ut i skilda grupper, så kallade *epistemiska kulturer* (Knorr Cetina 1999). För att kunna undersöka studiens frågeställningar om dataartikelns betydelse som en del av vetenskaplig kommunikation har jag valt en kvalitativ ansats med intervjuer som metod, något som motiveras nedan. I detta kapitel beskriver jag också de forskningsetiska hänsynstagande som gjorts och de analytiska ansatserna. Slutligen reflekterar jag, med anledning av ämnet för uppsatsen, över min egen hantering av forskningsdata.

4.1 Intervjuer

Valet av intervju som metod föranleddes av att syftet var att undersöka vad som motiverar forskare att dela sina data och vilken funktion dataartikeln kan fylla i deras arbete och som en del av meriteringssystemet. Intervjuer ger möjlighet att ta del av informanternas reflektioner och beskrivningar av erfarenheter runt delning och publicering av forskningsdata, något som hade varit svårt att göra genom exempelvis enkäter (Eriksson-Zetterquist & Ahrne 2015). Som tidigare nämnts i avsnittet om kvalitativa och kvantitativa data (se avsnitt 2.1.3), har intervjuer kritiserats för att inte säga så mycket om ”den sociala verkligheten utanför intervjusituationen” (Alvesson 2011, s.). Att närma sig forskarnas utikispunkt (Lalander 2015) genom att komplettera intervjuerna med observationer på plats i deras arbetsmiljö hade därför varit önskvärt. Just de handlingar som utförs när en dataartikel skrivs hade dock varit svåra att studera i det här fallet eftersom relativt få svenska forskare skriver dataartiklar, och det dessutom hade varit nödvändigt att hitta dem innan de var färdiga med arbetet. Ett annat sätt att fördjupa bilden hade kunnat vara att få se informanternas data och hur de arbetar med, strukturerar, lagrar och delar dessa. Detta hade kunnat uppnås antingen genom observationer eller genom att utföra alla intervjuer på plats i intervjupersonernas arbetsrum, men eftersom forskarna var utspridda över landet var detta tyvärr inte heller möjligt inom ramen för denna uppsats.

4.1.1 Urval

Genom sökningar i Web of Science Core Collection och i Scopus¹¹ hittade jag sammanlagt trettio forskare knutna till svenska lärosäten som hade skrivit och/eller citerat minst

¹¹ I Web of Science fanns möjligheten att välja data paper som artikeltyp, vilket underlättade sökningarna. I Scopus fanns inte detta alternativ så i de sökningarna fick jag gå en omväg genom att söka efter namngivna data journals och i nästa steg efter författare som publicerat sig i dessa data journals och som var knutna till svenska lärosäten.

en dataartikel och som angivits som ”corresponding author” för den aktuella artikeln.¹² För en fördjupad förståelse av praktikerna runt delning av forskningsdata sökte jag även efter forskare som tillgängliggjort sina data i Svensk Nationell Datatjänsts databas och letade då särskilt efter forskare som delat kvalitativa data. Bland de forskare, som fortfarande var knutna till respektive lärosäte och på så sätt möjliga att kontakta via mejl, valde jag ut och skickade intervjuförfrågningar (se [bilaga 1](#)) till sjutton.

I urvalet strävade jag efter att täcka in så många olika forskningsområden som möjligt inom den relativt begränsade tidsram en uppsats har. Jag försökte även få en balans mellan forskare som hade delat kvalitativa och kvantitativa data. Detta hade flera skäl: Dels ansåg jag det vara intressant att genom intervjupersonernas utsagor försöka skapa mig en bild av hur dataartiklar används inom olika discipliner och forskningsprojekt med olika typer av data och se om det gick att urskilja likheter och skillnader. Dels var risken uppenbar att jag inte hade hittat tillräckligt många intervjupersoner om jag hade valt att begränsa undersökningen till ett visst forskningsområde eller en viss typ av data. Av de sjutton forskare jag kontaktade var det åtta som slutligen hade möjlighet att delta i undersökningen. De forskar inom sociologi, språkvetenskap, naturgeografi, ekosystemvetenskap, statsvetenskap, klinisk medicin, ekonomisk demografi och ekologi. Deras akademiska karriärer är olika långa: tre är doktorander; en har en postdok-tjänst; en är universitetslektor; en har doktorerat och arbetar sedan länge som externfinansierad forskare; och två är professorer sedan många år tillbaka. Det visade sig vid intervjuerna att två av de tre forskare som jag hittade genom Svensk Nationell Datatjänst (SND) – och som jag alltså hade tänkt intervjua utifrån de erfarenheter de hade av andra sätt att dela data – även hade varit medförfattare till dataartiklar som jag inte hade hittat under mina sökningar i Web of Science och Scopus. Därmed var det bara en intervjuperson som inte hade någon tidigare erfarenhet av dataartiklar. Sex av forskarna är medförfattare till en eller flera dataartiklar och av dessa har två även granskat andras dataartiklar. En forskare har enbart citerat dataartiklar.

Frågan om hur många personer som ska intervjuas för att uppnå generaliserbarhet och trovärdighet är svår och måste även anpassas till den utsatta tidsramen. Ulla Eriksson-Zetterquist och Göran Ahrne (2015) beskriver betydelsen av att uppnå mättnad för att representativiteten i kvalitativa intervjuer ska kunna diskuteras: ”När studenten eller forskaren upplever att man känner igen svaren och att samma svarsmönster återkommer i ett flertal intervjuer kan man tala om en sådan mättnad” (ibid., s. 42). Genom att varva intervjuer med analys går det att avgöra när mättnad uppnåtts och ytterligare intervjuer inte längre skulle ge ny kunskap. Ett annat sätt att öka studiens generaliserbarhet är genom att sätta den i relation till mer generella begrepp eller teorier (Svensson och Ahrne 2015), något jag strävar efter att göra i analysdelen (se kapitel 6).

4.1.2 Genomförande av intervjuerna

Uppsatsen bygger huvudsakligen på kvalitativa intervjuer, löst strukturerade efter två olika intervjuguider (se [bilaga 2](#)): en för de forskare som har skrivit dataartiklar och en för dem som enbart har citerat dataartiklar eller som har delat data på andra sätt. Intervju-

¹² Många dataartiklar hade tio eller fler författare och jag gjorde bedömningen att det inte var intressant att intervjua forskare som stod längre ned på listan över medförfattare eftersom det fanns en risk att dessa bara hade bidragit med data och inte varit delaktiga i författandet av själva dataartikeln.

guiderna är uppdelade i fyra teman som rör 1) grundläggande fakta om informantens data; 2) dataartiklar; 3) betydelsen av att dataartiklar kan granskas, citeras och bidra till akademisk meritering på ett liknande sätt som en konventionell artikel; och 4) praktiker runt delning av forskningsdata. Intervjufrågorna handlade huvudsakligen om forskarnas erfarenheter av publicering och delning av data och skillnaderna mellan olika forskningsområden hamnade till viss del i bakgrunden. Materialet samlades in under februari och mars 2018.

Fyra av intervjuerna genomfördes via videosamtal, två gjordes på plats i forskarens egen miljö och två av forskarna bad att få svara på frågorna via e-post på grund av tidsbrist. Detta innebär att omständigheterna för intervjuerna varierade. Framförallt bör påpekas att de forskare som svarade via mejl fick samtliga frågor tillskickade sig vid samma tillfälle. Det var därmed inte fråga om regelrätta e-postintervjuer där en fråga i taget ställs och varje svar inväntas innan nästa fråga skickas (Kvale & Brinkmann 2009). Trots att uppföljande frågor i båda fallen ställdes och besvarades via mejl fanns inte samma möjligheter till fördjupande följdfrågor och förtydliganden. Vid dessa tillfällen följdes alltså frågeguiderna strikt på ett sätt som snarare liknar ”strukturerade intervjuer med standardiserade frågeformulär” (Eriksson-Zetterquist & Ahrne 2015, ss. 35) eller enkätundersökningar med öppna intervjufrågor (ibid.). Eftersom mitt huvudsakliga syfte inte var att jämföra intervjupersonernas svar gjorde jag bedömningen att det var relevant att inkludera även dessa svar. Övriga intervjuer var löst strukturerade utifrån frågeguiderna: samtliga teman täcktes in, men frågorna, den ordning de ställdes i och, i förekommande fall, följdfrågorna anpassades efter situationen. I detta avseende var det ingen skillnad mellan de intervjuer som genomfördes vid videosamtal och de intervjuer som gjordes på plats. Däremot kan det förstås vara svårare att avläsa nyanser och reaktioner vid ett videosamtal (Eriksson-Zetterquist & Ahrne 2015), men då intervjuerna inte berörde personliga eller känsliga ämnen bedömer jag att detta inte har haft någon avgörande betydelse för utfallet. Med informanternas tillåtelse gjordes en ljudinspelning av intervjuerna, som minnesstöd och för att möjliggöra citering. De intervjuer som genomfördes vid direkta möten eller videosamtal har blivit utskrivna i sin helhet, men inte på en sådan detaljnivå att exempelvis harklingar och upprepningar av småord tagits med. När informanterna citeras i resultatdelen har språket i vissa fall justerats lite ”för att det ska bli läsvänligt och för att den intervjuade ska framställas på ett rättvist sätt” (Eriksson-Zetterquist & Ahrne 2015, s. 51). Detta gäller även de svar som kom in via mejl, där exempelvis förkortningar har skrivits ut för att ge ett bättre flyt i texten. Närmare omständigheter runt intervjuerna redovisas i ett appendix, tillsammans med forskarnas yrkestitel och forskningsområde. Där kopplas även citaten till de olika intervjuerna.

4.1.3 Forskningsetiska överväganden

Alla intervjuer genomfördes med intervjupersonernas informerade samtycke (se [bilaga 3](#)). De deltagande forskarna informerades om studiens syfte och att deltagande var frivilligt och kunde avbrytas när som helst och utan förklaring. Av det informerade samtycket framkommer att forskarnas namn inte kommer att anges i uppsatsen och inte heller vid vilket lärosäte de arbetar. Däremot framgår det vilken befattning intervjupersonerna har (till exempel professor, docent, doktorand), och vilka forskningsområden de är verksamma inom. Detta kan, tillsammans med det faktum att så få forskare i Sverige har publicerat dataartiklar, göra det svårt att uppnå fullständig konfidentialitet, något som de intervjuade accepterade när de skrev under det informerade samtycket. Intervjumaterial som samlats in i studien har avidentifierats och forskarnas namn har ersatts med koder (se

appendix). Kodnyckeln bevaras så att obehöriga inte kan komma åt den, och på annat ställe än insamlade data (Trost 2014). I vissa citat har detaljuppgifter, som inte är betydelsefulla för sammanhanget men som skulle kunna bidra till att avslöja forskarnas identitet, tagits bort. I det material som jag har bedömt vara mest känsligt, och som huvudsakligen gäller synen på forskarkollegor som inte delar data, har jag valt att inte koppla citaten till enskilda intervjupersoner. Jag vill särskilt betona att detta inte är något som intervjupersonerna själva har föreslagit eller krävt utan något jag har valt att göra för att minimera risken för negativa konsekvenser för deltagarna. Även dessa citat är kopplade till koderna, men informationen bevaras tillsammans med kodnyckeln (se ovan). De intervjuade forskarna presenteras i inledningen av kapitel 5 och sedan följer resultatet av undersökningen.

4.2 Presentation av resultatet och analysens genomförande

Analys av kvalitativt material är en process som involverar flera olika delmoment. Dessa moment har beskrivits som att sortera, reducera och argumentera (Rennstam & Wästerfors 2015) eller som att identifiera teman och underteman, sälla ut ett hanterbart antal samt länka samman teoretiskt ramverk och teman (Ryan & Bernard 2003). I artikeln "Techniques to Identify Themes" (Ryan & Bernard 2003) beskrivs hur både det empiriska materialet i sig och tidigare (a priori) teoretisk förståelse av ett fenomen kan ge upphov till teman. Eftersom frågeställningarna är utforskande till sin natur har jag strävat efter att närma mig intervjumaterialet så förutsättningslöst som möjligt. Visserligen är det mer eller mindre oundvikligt att val av teman och frågor i intervjuguiderna, som ju kan sägas utgöra en allra första sortering av materialet, har präglats av min förförståelse av ämnet. Samtidigt är det viktigt att inte i analysen låsa sig vid dessa val då det är omöjligt att på förhand förutse alla de teman som kan uppstå (Ryan & Bernard 2003 se Dey 1993, ss. 97-98). Även när materialet reduceras är det naturligtvis viktigt att inte välja bort det som inte passar in i den förväntade bilden utan att sträva efter representativitet (Rennstam & Wästerfors 2015).

För att underlätta sortering av materialet användes analysprogrammet Nvivo. Samtliga intervjuer, både i form av transkriptioner och mejlsvar, samlades i ett dokument för att kunna analyseras som en helhet. Därefter kodades intervjuvären i en initial sortering där jag inte bara sökte efter likheter utan även efter skillnader och utsagor som bröt mönstren (Rennstam & Wästerfors 2015; Ryan och Bernard 2003). De teman som identifierades under denna första sortering visade sig i stor utsträckning sammanfalla med just de frågor som Christine Borgman (2012, se inledningen av kapitel 2) föreslår som underlag för att bättre förstå praktiker och underbygga policyer för datadelning: "which data might be shared, by whom, with whom, under what conditions, why, and to what effects" (ibid., s. 1072). Dessa frågor fick därför bli vägledande för att strukturera och presentera intervjuresultaten under olika teman (se kapitel 5). Inspirerad av Jutta Haider och Sara Kjellberg (2016; se även avsnitt 2.1.2) valde jag att lägga till en fråga om den tidsmässiga aspekten av datadelning. Vissa teman är mindre omfattande och har därför slagits ihop i samma avsnitt, medan andra, som kräver mer utförliga svar, har fått ett eget avsnitt. Intervjumaterialet presenteras under följande rubriker: "*Vem delar data och med vilka användare i åtanke?*", "*Vilka data delas och vilka delas inte?*", "*När delas data?*", "*Varför delas data?*", "*Hur går det till att dela data?*" samt "*Vad innebär det att dela data?*". Trots att uppsatsens frågeställningar om dataartikeln egentligen handlar om att *publicera* data (se avsnitt 2.3) har jag valt att behålla formuleringen *dela* från Borgmans citat, och använder

det i resultatkapitlet som ett paraplybegrepp som kan omfatta både att tillgängliggöra ett dataset och att beskriva sådana data i en dataartikel. Detta eftersom intervjuerna visade att de båda aktiviteterna är så tätt sammankopplade att de kan vara svåra att separera. Dessutom tillåter ordvalet att även svar från de forskare som inte har skrivit dataartiklar inkluderas, liksom beskrivningar och jämförelser som intervjupersonerna gjorde av andra sätt att dela data.

I nästa steg kodades intervju svaren på nytt för att möjliggöra en analys där det empiriska materialet länkades samman med det teoretiska ramverket och även med tidigare forskning. Ryan och Bernard (2003) föreslår en metod för att hitta teorirelaterade teman där materialet söks igenom efter utsagor som belyser frågor som är centrala i samhällsvetenskaplig forskning. De kan till exempel röra metoder att uppnå status eller social kontroll, tecken på social konflikt, information om hur problem löses, men även hur informanter uttalar sig om "människor, objekt, processer, aktiviteter, händelser och relationer" (ibid., s. 93 se Bogdan & Biklen 1982, ss. 156-162, egen översättning). En fara med att allt för mycket låsa fast sig vid de teoretiska utgångspunkterna är, varnar författarna, att forskaren bara hittar det han eller hon letar efter. Att å andra sidan helt undvika teorierna skapar i stället en risk att missa kopplingar mellan data och viktiga forskningsfrågor (Ryan & Bernard 2003). Även här gällde det alltså att närma sig materialet med så öppen blick som möjligt. Tre övergripande teman identifierades som centrala i forskarnas beskrivningar av praktikerna runt dataartiklar och delning av data: "*Förtroende och trovärdighet*", "*Samarbete och erkännande*" samt "*Påverkan och status*". Dessa diskuteras i kapitel 6.

4.3 Egen hantering av forskningsdata

Eftersom jag inte är universitetsanställd och min studie inte är finansierad av offentliga medel finns det inga krav på hur jag hanterar mina forskningsdata, men uppsatsens ämne har förstås gett mig anledning att fundera på frågan vid upprepade tillfällen. De resurser för forskare som HT-biblioteken vid Lunds universitet har samlat på sin webbsajt under rubriken Forskningsdatahantering¹³ blev en utgångspunkt i dessa funderingar. Där finns råd om datahantering inför, under och mot slutet av ett forskningsprojekt. Steg ett är att skapa en datahanteringsplan för att säkerställa att insamlade data hanteras på ett effektivt sätt under projektet och för att underlätta arbetet med att förbereda dessa data för arkivering, delning och framtida användning. En resurs biblioteket hänvisar till är Svensk Nationell Datatjänsts checklista för datahanteringsplaner¹⁴ som är indelad enligt följande rubriker:

- **Översikt:** Namn på projekt, medverkande forskare och huvudman (dvs ägare av de data som produceras), roller, finansiärer och riktlinjer (t ex policy för IT-säkerhet).
- **Skydda dina data:** Frågor som ställs under denna punkt är om det finns behov av etikprövning, informerat samtycke, skydd av respondenternas identitet eller om personuppgiftsombud behöver kopplas in, samt om projektet omfattar konfidentiell information eller om det finns upphovsrättsliga frågor att ta hänsyn till.

¹³ <http://www.htbibl.lu.se/forskare/forskningsdatahantering/>

¹⁴ https://snd.gu.se/sites/snd.gu.se/files/Checklista%20datahanteringsplan_2017-12-12.pdf

- **Samla in/producera data:** Här beskrivs redan befintliga data och data som ska samlas in och med vilken metod och vilka resurser som krävs.
- **Dokumentera dina data:** Hur ska informationen struktureras och föras med metadata så att alla inblandade i projektet och framtida användare förstår det? Finns det metadatastandarder eller etablerade ontologier?
- **Organisera dina data:** Tar upp frågor om hur materialet ska struktureras med hjälp av mappstruktur och filnamn och vilket/vilka filformat ska användas. Hur ska det lagras och säkerhetskopieras?
- **Kostnader för hantering av data:** Vad kostar personal, hård- och mjukvara och lagring av data?
- **Bevara och tillgängliggör dina data:** Hur kan data och dokumentation förberedas för tillgängliggörande i enlighet med arkivlagen? Finns det begränsningar i vilket material som kan tillgängliggöras, exempelvis av etiska eller juridiska skäl eller på grund av att det krävs någon särskild programvara. Andra frågor som ska besvaras här gäller upphovsrätt, önskemål om hur data ska citeras och materialet kommer att få en så kallad persistent identifier (PID). Kontakt med ett datarepositorium (t ex SND:s eget) krävs redan nu om tanken är att materialet ska tillgängliggöras.

Checklistan inleds med en ruta för datum och version. Tanken är att checklistan är ett levande dokument som kan behöva uppdateras flera gånger under ett forskningsprojekt. Den första punkten är inte relevant i mitt fall eftersom jag är ensam om projektet och inte har någon finansiär. När det gäller behov av att skydda data är det främst frågorna runt informerat samtycke och skydd av respondenter som väcker frågor. Av forskningsetiska hänsyn valde jag att aidentifiera intervjupersonerna, något som också framgick av det informerade samtycket (bilaga 2). Att redigera bort alla delar av intervjuerna som innehåller uppgifter som kan bidra till att intervjupersonerna kan identifieras skulle inte bara vara enormt tidskrävande – frågan är hur mycket av materialet som skulle vara kvar efteråt, och hur osammanhängande det skulle vara. För att kunna tillgängliggöra intervjuerna hade blanketten för informerat samtycke också fått omformuleras så att respondenterna godkände all slags framtida användning av materialet utan att känna till sammanhanget. Frågan är om detta hade påverkat deras vilja att bli intervjuade. Redan i nuläget var det svårt att hitta forskare som hade tid och ville svara på mina frågor. Flera av de följande punkterna på checklistan ligger ganska nära det planeringsarbete som ändå bör göras innan ett projekt dras igång. Att beskriva vilka data som ska samlas in och med vilka metoder är till exempel något som bör finnas i metodavsnittet så det innebär inget extraarbete. Hur data ska struktureras i mappar, vilka filformat som ska användas och hur det ska säkerhetskopieras är frågor som är bra att fundera över oavsett om data ska tillgängliggöras eller inte. Att lägga till metadata och på andra sätt förbereda sina data för långtidsbevaring är däremot något som bara är värt att ägna tid åt om de verkligen ska tillgängliggöras.

Redan för ett så litet projekt som en uppsats inbegriper alltså upprättandet av en datahanteringsplan många tidskrävande moment. Troligen är det så att ju större datamängder ett projekt genererar desto större nytta kan en datahanteringsplan göra. Från start synliggörs då vilken infrastruktur som krävs samt tidsåtgång och kostnad för datahantering, och när forskningsansökan skickas in är det troligen lättare att söka särskilda medel för detta. Om planen är att tillgängliggöra forskningsdata finns det säkert mycket tid att vinna på att redan från början förbereda sina data för delning.

En intressant fråga som inte tas upp i SND:s checklista är den om vilka data som är värda att tillgängliggöra. I mitt fall utgör intervjuerna visserligen ett unikt material, men hur stor är chansen att någon skulle vilja återanvända det? Det är knappast ekonomiskt hållbart att lägga mycket tid på att dela och bevara data till ingen nytta. Samtidigt är det, som tidigare har framhållits, mycket svårt att i förväg veta vilka data som kan bli användbara i framtiden. En annan reflektion är att delning av data, i alla fall när det som här rör sig om intervjuer, inte bara exponerar intervjupersonerna utan även forskaren själv. Exempelvis lämnas brister i intervjuteknik, som dåligt formulerade frågor eller luckor i intervjun som kanske måste kompletteras i efterhand, ut öppet. När Christine Borgman besökte Lunds universitet som huvudtalare vid ett symposium¹⁵ fick jag möjlighet att ta upp bland annat denna fråga med henne och hon konstaterade att det inte är ovanligt att forskare tror att alla andras data är ”prettier than your own”. Frågan är hur stor betydelse en sådan känsla kan ha när beslut fattas om vilka data som ska delas och inte?

¹⁵ Symposium som en del av Theme DATA, Lunds universitet den 3 maj 2018.

5. Intervjuerna med forskarna

Resultaten från intervjustudien presenteras här under sex teman som framkom när materialet kodades i ett första steg. Dessa teman rör vem som delar data, vilka data som delas, när och varför data delas, hur det går till och vad det innebär. Samtliga intervjupersoner forskar vid svenska universitet och har skrivit eller citerat minst en dataartikel och/eller aktivt delat sina data på andra sätt. Deras ställning och erfarenheter av datadelning presenteras här kortfattat:

1. Professor. Har delat data i SND:s databas.
2. Externfinansierad forskare (har doktorerat). Medförfattare till en dataartikel som inte hade publicerats vid intervjutillfället. Har även delat data i SND:s databas.
3. Universitetslektor. Har skrivit två dataartiklar och håller på att slutföra en tredje.
4. Doktorand. Har skrivit en dataartikel.
5. Professor. Har skrivit, citerat och granskat flera dataartiklar samt delat data i SND:s databas.
6. Doktorand. Har skrivit en dataartikel.
7. Doktorand. Har citerat en dataartikel.
8. Forskare med postdok-tjänst. Har skrivit en dataartikel.

Uppgifter om omständigheterna runt intervjuerna återfinns efter källförteckningen (se appendix), och där kopplas citaten också samman med respektive intervju i slutfotnoter.

5.1 Vem delar data – och med vilka användare i åtanke?

Gemensamt för de tre forskare som har varit verksamma längst är att de verkar inom forskningsfält där det sedan flera decennier tillbaka finns både en stark tradition av och funktionella vägar att dela data. Två av dem har deltagit i större projekt med forskare från flera länder där kvantitativa data har samlats in och delats i gemensamma databaser, bland annat för att möjliggöra jämförelser. ”Jag kom in i [projektnamn] som doktorand, och det var fullkomligt naturligt att släppa alla data fritt. Det handlar mycket om vad du socialiseras in i”.¹ Den tredje forskaren, som har doktorerat och jobbar som externfinansierad forskare sedan flera år tillbaka, arbetar huvudsakligen med kvalitativa data i mindre projekt. Här är det traditionen inom ämnet, snarare än inställningen hos de närmaste kollegorna, som har gjort det till en självklarhet att dela data.² När de etablerade kanalerna för att dela data inte har fungerat av någon anledning – till exempel för att det har funnits extramaterial som inte har passerat in i det större projektet eller kvalitativa data från egna, mindre projekt – har forskarna aktivt sökt andra sätt att dela dessa data. De har då bland annat använt sig av SND:s databas. Den ena forskaren säger sig aldrig ha haft behov av några andra sätt att dela data.³ Den forskare som enbart hade citerat, men inte själv skrivit, en dataartikel delar i stället sin data ”som supplement till artiklarna på tid-

skriftens hemsida” och planerar även att ladda upp sin data till en större forskningsdatabas online så snart den ”senaste artikel[n] fått ett DOI”.⁴

Flera av de forskare som har varit verksamma kortast tid har inte alls samma bild av en heterogen tradition inom sitt forskningsfält, utan beskriver i stället hur vissa, oftast äldre, forskare är snåla med sina data, men att det bland yngre forskare är närmast självklart att data ska delas och att kulturen runt delning av data förändras snabbt i denna riktning.⁵

Nu är det ju så etablerat och accepterat att man ska dela med sig av sin data så att de som inte delar sin data får faktiskt skämmas lite för det. Nu är ... normen är att dela med sig och alla håller med om nu att det är lite gammaldags och fult att inte göra det.⁶

Att inte dela med sig av sina data beskrivs som ett omodernt arbetssätt som bara de kan tillåta sig som har jobbat länge inom ett fält och ”vet att de är tillräckligt kända och väletablerade så att de får fortsatt finansiering ändå”.⁷ En sådan attityd fungerar inte för den som vill göra akademisk karriär idag, enligt flera forskare. En intervjuperson betonar vikten av att inte få ett dåligt rykte: ”Det är viktigt för en forskare hur folk pratar om en när de sitter och tar en öl på en konferens, lite off-the-record, innan ... Alltså det är otroligt viktigt”.⁸ Att inte vara generös med sina data riskerar alltså att förstöra framtida samarbetsmöjligheter.

Folk som är framgångsrika tenderar, mer eller mindre, inom givna ramar att kunna välja ungefär vilka man vill samarbeta med. Och då väljer man inte att samarbeta med folk som är krångliga och inte vill dela med sig av data, /.../ och inte ska lita på nån och fram och tillbaka. Det är liksom inte roligt. Det spelar ganska stor roll faktiskt.⁹

Forskarkollegor som kräver att få bli medförfattare till en artikel för att dela med sig av sina data är ett ämne som tas upp i fem av intervjuerna: ”Det finns vissa grupper, vissa forskningsledare för stora grupper, som fortfarande är så kallade *data huggers* och bara inte vill dela med sig av sina data, och aldrig delar med sig om de inte får vara medförfattare”.¹⁰ Ett sätt detta görs på är att ”de ser till att det endast är de som förstår hur datasetet är strukturerat”.¹¹ Att kräva medförfattarskap för att dela data beskrivs både som ”lite väl magstarkt”¹², något ”det har gått inflation i”¹³ och ”något som vi måste komma bort ifrån”.¹⁴

[D]en möjligheten försvinner ju lite nu. Nu får man en citering vilket jag kan tycka att det är väl mer fair, för det känns inte som att man ska ha ett medförfattarskap. Ett medförfattarskap har nog varit ganska lätt, det är lätt att slänga med en extra, men de gör ju oftast ingenting då, utan bara ger data.¹⁵

En forskare säger att det inom dennes disciplin till och med är så att många anser att ett sådant medförfattarskap är oetiskt och att ett kriterium bör vara, inte bara att man har producerat data eller ägt laboratoriet, ”utan att man faktiskt varit med och skrivit på artikeln eller gjort analyser som ligger till grund för artikeln”.¹⁶ Här skulle dataartikeln kunna bidra till att mer data görs öppet tillgängliga: ”Om de istället kan publicera dessa data i renommerade datatidskrifter som ger citeringar, borde delningen av data öka. Sedan tar det också tid att iordningställa data så att de kan användas och förstås av andra forskare, och där kan återigen dataartiklar motivera till delning av data”.¹⁷

Att intervjupersonerna hoppas att andra forskare ska kunna ha glädje av dataartiklarna och/eller den data som delas är nästan en självklarhet och gäller oavsett vilken typ av data

som delas. Flera arbetar med stora, internationella projekt där utbytet av material så att säga är inbyggt från början: ”Mycket av datan kan ses som en form av miljöövervakning där forskarkollektivet hjälps åt att kartlägga vissa ekologiska aspekter eller parametrar”.¹⁸ Men det kan tvärtom också handla om ett ämne där det finns så lite data att det blir viktigt att tillgängliggöra den för kollegor.¹⁹ Myndigheter är en annan tänkt användare, liksom relaterade yrkesgrupper. I en medicinrelaterad öppen databas, som en av forskarna i samarbete med kollegor beskrivit i en dataartikel, gjordes en enkät för att ta reda på mer om vilka som använder resursen: ”De flesta var faktiskt apotekare eller farmaceuter [...] sedan var det också en del ST-läkare och sjuksköterskor, så det visar ju på att det är olika professioner som använder det”.²⁰ Några av projekten har som uttalat syfte att data ska delas även med allmänheten.²¹ ”[D]et är spännande att det är öppet för så många olika grupper, det är ju det som är målet med open data att det ska vara öppet även för allmänheten”.²² Och i ett fall nämns särskilt syftet att göra insamlade data, som omfattar dokumentation av ett hotat minoritetsspråk, tillgängliga för informanterna själva: ”[S]å de får se sitt eget språk [...] och kan få lite hjälp på vägen om de själva vill bevara det, så går det att göra”.²³

Frågan om vilka användare data delas med berör också tillgänglighetsaspekten: Vem har kännedom om att dessa data finns och hur ges tillgång till dem? Informella sätt att dela data förekommer fortfarande. Det kan röra sig om data som forskaren ännu inte har hunnit förbereda för att de ska kunna delas officiellt.

Skilnaden där är att jag oftare beskrivit dessa data muntligt eller i ett email, och inte lagt lika mycket tid på att strukturera och beskriva data så att de kan förstås utan min hjälp. Ibland har jag också skräddarsytt vissa data specifikt för kollegans ändamål.²⁴

En forskare berättar om sin första dataartikel att den beskriver en databas som forskaren tagit över från en kollega och uppdaterat: ”Innan dess hade den inte funnits tillgänglig, utan han skickade ut en cd-rom till någon som hörde av sig ibland”.²⁵

5.2 Vilka data delas och vilka delas inte?

Som tidigare nämnts var ambitionen att hitta intervjupersoner från flera olika ämnesområden och detta avspeglar sig också i vilka sorters data som intervjupersonerna har delat (och sedan i flera fall beskrivit i dataartiklar). Det rör sig om både kvalitativa och kvantitativa data i form av bland annat ljudinspelningar, bilder med metadata och transkriptioner, geografiska data i så kallade shapefiler¹⁶, fotografier av historiska dokument, en geospatial databas, statistik och dokumentation av offentlig kommunikation. Som redan nämnts skiljer sig dataseten också åt i omfattning såtillvida att det både rör sig om avgränsade studier där en eller ett par forskare deltagit, och enorma projekt där data samlats in av forskargrupper över hela världen och sedan sammanställts. Flera forskare säger sig dela alla data – ”som princip”²⁶ och så fort tid medges.

[J]ag har svårt att komma på exempel på data som jag har samlat in som jag inte sedan har delat. Då får jag gå ända tillbaka till avhandlingen, då gjorde jag fallstudier, kvalitativa intervjuer [...]. Dem har jag aldrig delat, de ligger fortfarande på kassettband i källaren.²⁷

¹⁶ Shapefil, ett vanligt format för geografiska data.

Öppenhet kopplad till transparens och vetenskaplig akribi är en central princip som kan väga tungt och påverka utformning av metoder, till exempel vid kvalitativa intervjuer. Ibland måste en avvägning göras mellan valet att låta intervjupersoner vara helt anonyma och på så sätt ibland kunna "få väldigt fullständig information"²⁸ eller att vara helt öppen och koppla citat till namngivna informanter: "Jag höll väldigt strikt på att varje sak som jag skrev gick att koppla explicit till intervjun [...] Och det är klart att ibland betydde det att jag fick mindre information än vad jag annars hade kunnat få. Så jag premierade det kriteriet".²⁹ Två av de forskare som delat data i SND:s databas kommenterade det faktum att det där går att välja tillgänglighetsnivå så att forskarens godkännande måste inhämtas innan data delas.¹⁷

En fördel med dataartiklar som en av forskarna tar upp är att de kan beskriva även negativa eller neutrala utfall i studier, som forskare kanske inte vill lägga utrymme på att beskriva i en konventionell artikel: "Om det blir nåt negativt utfall till exempel då tänker man att det här är inte värt att publicera".³⁰ Sådana uppgifter kan bli värdefulla i andra sammanhang, betonar forskaren som har arbetat med en sammanställning av könsskillnader i medicinska sammanhang. I cirka en tredjedel av de studier som sammanställts går det inte att veta om studierna har gjorts uppdelade efter kön. "Förmodligen har de det, men de presenterar det inte. Och även om det kanske inte finns könsskillnader, så kan det ju vara värt att veta att de har studerat det och bedömt att här finns inget att säga".³¹

Tre av intervjupersonerna har tagit på sig arbetet att sammanställa andras data i större databaser: "Det började komma väldigt mycket data [för tio år sedan], men den var helt utspridd, så nån hade sin data på en hemsida här och nån annan på en annan hemsida där och vi bara tog på oss att sammanställa detta".³² Att samla data från många olika källor kan innebära att mycket arbete får läggas på konvertering. Även om det finns standarder är det inte alla som följer dem så noga och "alla grupper har ändå lite olika sätt att lagra data".³³ Samma forskare berättar att deras forskargrupp numera uppmanar kollegor att skicka in sina rådata som de är:

Vi går från normen att folk ska skicka in data efter en viss mall till att säga att om ni har tillräckligt mycket data så skriver vår programmerare ett script som översätter er data till rätt format, och så lagrar vi allt i ett format. Vi tror att det är vägen framåt för att få folk att dela med sig av data.³⁴

I de fall intervjupersonerna *inte* delar alla sina data har det ofta rört sig om integritetsskäl eller känsliga data som är skyddade av personuppgiftslagen¹⁸ och som "därför inte får delas hur som helst".³⁵ Som exempel nämns ett projekt om lokalpolitiker: "Det törs vi inte tillgängliggöra, det är för enkelt att identifiera dem. Men det är inte av egoistiska skäl vi inte delar det".³⁶ Forskaren nämner dock möjligheten att tillgängliggöra materialet i ett repository, men med restriktioner för vem som får använda det och hur. Ett annat exempel gäller ett projekt med både historiska och moderna data på individnivå: "[N]är jag har delat relaterade geografiska data har detta endast varit sådana data som inte kan kopplas till moderna individdata som är skyddade av bland annat personuppgiftslagen".³⁷ En forskare konstaterar att "om man har samlat in data själv då gäller det förstås att ha koll på sina kodningar så att man inte råkar dela data som är inte är anonymiserad".³⁸ Men

¹⁷ Se Svensk nationell datatjänsts webbsajt. <https://snd.gu.se/sv/beskriv-och-lamna-in-data/tillganglighetsnivaer>

¹⁸ Den 25 maj 2018 ersattes Personuppgiftslagen av Dataskyddsförordningen (se avsnitt 2.2.2).

även i de fall materialet innehåller känsliga uppgifter räcker det dock ibland att informanternas identitet koddas eller att vissa delar tas bort innan materialet delas:

Det är ju inte om hemliga saker, det är om namngivning, det är om ritualer och seder och förändringar och så, hur man tycker ja, om förändringar på sikt. Och hur det var förut också. Så det ... ja, det handlar inte om politik. Jag har tagit bort alla sådana ... ja, så det finns inget sådant där.³⁹

Några forskare tar också upp att det kan vara så att en del forskare är osäkra på hur det fungerar med upphovsrätt och inte vet ”om man äger sin data”.⁴⁰ Som exempel nämns att det kan finnas en policy som säger att det är arbetsgivaren som har upphovsrätt till data och inte den enskilde forskaren. Detta gäller särskilt när andras data sammanställts och använts för egna analyser. Det kan också uppstå upphovsrättsliga konflikter i ”stora forskarlag eller samarbeten som ska komma överens om vad som slutligen ska göras med genererad data”.⁴¹ Men det är inte bara frågetecken runt upphovsrätt som gör att sådana data inte delas, utan även det faktum att det inte är egen ”unik rådata”⁴² utan data som redan finns öppet tillgängliga för alla att söka rätt på (se även under nästa avsnitt om när data ”blir till”). En forskare tror också att ”viss data inte delas av den enkla anledningen att man inte är medveten om att någon annan kan ha nytta av den”.⁴³ Samma intervjuperson tror att forskare kan vara återhållsamma med data eftersom deras karriärer hänger på den forskning – och de data – de producerar. ”Det går tyvärr inte alltid att lita på att information som delas öppet används på ett forskningsetiskt korrekt sätt eller ger cred till rätt person (i form av citeringar)”.⁴⁴

5.3 När delas data?

På frågan om det finns en viss punkt i arbetet där arbetsmaterial blir data varierar svaren. En forskare svarar nej utan att tveka: Det är data från början”.⁴⁵ En annan säger att det är när man ”känner sig färdig med det på något sätt – liksom, nu är det här klart och jag kan se att andra kan använda det”.⁴⁶ Enligt samma forskare är ett annat kriterium för att något ska ses som data som är värdefulla att dela att det måste ha skett något med dem, till exempel att olika uppgifter kombineras eller nya tillförs. ”[O]m man, som ett exempel, tar ett dataset med temperaturmätningar från en plats och så gör man ett medelvärde av det, då har det ju inte blivit ett dataset av det”.⁴⁷

Ett skäl att vänta med att dela data är hänsyn till juniora forskares karriärer: ”[D]e som samlar in rådatan är oftast doktorander och där är det så känsligt med publiceringar, så då tenderar vi att vilja att de ska ha fått ut sina artiklar innan vi öppnar upp datan”.⁴⁸

Data delas ofta i slutet av projekt och det kan göra det ännu svårare att få tiden att räcka till eftersom ett nytt projekt ofta ska dras igång samtidigt:

Jag är externfinansierad helt och hållet och jag har ju haft förmånen att gå från ett projekt till ett annat direkt och /.../ då är man väldigt otålig om man är igång med nästa redan ... Det tog fasansfullt mycket tid [att dela data] och man blir så otålig för det är ju när man avslutat ett projekt och då vill man inte hålla på med det, för då har man något nytt på gång.⁴⁹

I ett fall publicerades dataartikeln och den applicerade artikeln¹⁹ parallellt vilket beskrevs som något av en chansning: Om dataartikeln inte hade blivit antagen hade hela metodbeskrivningen fått flyttas till den applicerade artikeln – ”som redan var ganska omfattande”.⁵⁰ Att istället kunna referera till dataartikelns beskrivning av hur datasetet tagits fram var då en fördel. En annan forskare valde att publicera en av sina dataartiklar redan före den applicerade artikeln.⁵¹ I det fallet gavs förklaringen att dataartikeln byggde på en uppdatering av material som redan hade publicerats och att ”det var läge att få ut den”.⁵² Nästa dataartikel hade dock forskaren tänkt sig publicera parallellt med den applicerade artikeln.

Publiceringen av en dataartikel kan också försenas på grund av att peer review-processen drar ut på tiden: ”Den här peer review-processen har tydligen tagit så lång tid för det är ingen som har kunnat bedöma den”.⁵³ Enligt en av intervjupersonerna är det ett problem att det finns så få datatidskrifter inom det aktuella forskningsområdet och det gör att det kan ta väldigt lång tid att få in en dataartikel.⁵⁴

Flera av forskarna nämner tidsbrist som en orsak till att data inte delas: Det tar tid att strukturera och beskriva data innan de kan delas.

Den där tiden som går från att man har sitt eget ful-Excelark till [...] att få det tolkningsbart och fungerande för en annan, det innebär ganska mycket jobb. Och det där, det är ingen som betalar för det. Det är klart att det största problemet med att vi inte får mer data öppen här, det är att folk inte hinner, inte att de inte vill.⁵⁵

När delning av data dröjer kan det också bero på att kvalitetssäkringen är tidskrävande. En forskare berättar om ett projekt där finansiärerna krävde att data skulle delas direkt när projektet var klart: ”Det vanligaste för oss är ju att de stora forskningsråden finansierar, Riksbankens Jubileumsfond eller nåt sånt och så får man hålla på i fem sex år och sen publicerar man kanske sina data”.⁵⁶ I det här projektet hade de cirka två år på sig och kvalitetssäkringen var långt ifrån klar när de var tvungna att dela sina data.

Det var fruktansvärt stressigt. Det var några nätter när man inte sov så bra. Men samtidigt fantastiskt stimulerande för att det tvingade ju igång en typ av organisation som nästan liknar ... Jag kan tänka mig när Apple ska komma med en ny typ av Iphone och otroligt många människor jobbar mot samma mål för att få färdigt den här produkten.⁵⁷

Forskaren konstaterar att ”produkterna” i vanliga fall brukar vara de artiklar eller böcker som skrivs. ”Men här var det verkligen den här datan som var produkten. Det var så klart väldigt stressigt och skapade en hel del nervositet och frustration, men det var också väldigt kreativt och som ett enormt logistiskt lagarbete. Det var väldigt lärorikt”.⁵⁸

5.4 Varför delas data?

Gemensamt för de intervjupersoner som har skrivit dataartiklar eller på andra sätt aktivt delat sina forskningsdata är grundsynen att denna data inte är personlig egendom utan ”en kollektiv nytthet”⁵⁹ som fler ska ha möjlighet att använda. ”Detta gäller särskilt data

¹⁹ En artikel där data som beskrivs separat i en dataartikel används (se även ordlistan i avsnitt [1.4](#))

som har skapats inom skattefinansierade forskningsprojekt, men jag delar också data för att gynna forskningen i allmänhet”⁶⁰, säger en annan forskare. En intervjuperson tycker att det fanns ett helt annat embargotänkande för tio år sedan och att det har blivit lättare och lättare att få tillgång till data:

Normen går hela tiden mot att detta är kollektiva nyttigheter och det är ju intressant på det sättet att det är inte självklart att man blir så himla gynnad av att ha lagt ner all den här tiden, för det är inget enkelt sätt till citeringar liksom. Men jag tycker ändå att det går åt det hållet.⁶¹

Ett annat skäl är, som tidigare nämnts, att det inom vissa smalare forskningsområden finns väldigt lite material att tillgå. En intervjuperson betonar också att det är viktigt att bidra till den norm som anses vara önskvärd: ”[V]i försöker pusha andra grupper att dela data, så då får man ju föregå med gott exempel”.⁶² En annan anser att det är viktigt att dela data för att öka transparensen runt forskningen: ”[R]ent] principiellt [tycker jag] att det ska vara så tydligt som möjligt vad man gör och vad man baserar sina konklusioner på också. Det är ju en annan dimension”.⁶³

De intervjuersoner som själva har skrivit dataartiklar är överlag positiva. ”[M]an får lite mer motivation att dela det [datasetet] för att man kan göra ett sånt här paper. Och det tycker jag är en vettig grej att man får lite tillbaka om man har gjort ett dataset och andra vill använda det”.⁶⁴ För flera forskare har det också inneburit en möjlighet att på ett utförligare sätt kunna beskriva metoder eller experiment som bedömts vara intressanta för andra.⁶⁵ En fördel som nämns i flera intervjuer är att den applicerade artikeln då inte heller behöver belastas med detaljerade metodbeskrivningar. En forskare tycker att en dataartikel är ett bra sätt att snabbt få ut ett dataset eftersom det inte behövs någon ”diskussion som tar mycket tid att skriva. /.../ Det var ju inget att diskutera, utan vi beskrev metoden och vi publicerade alla koder som användes för att ta fram [datasetet] på GitHub”.⁶⁶ Av samma skäl säger den forskare som själv enbart har citerat dataartiklar att det kan tyckas vara ”ett lättvindigt sätt att addera saker på den akademiska meritlistan, då jag antar att det skulle vara mindre arbete att endast publicera data”.⁶⁷ Forskaren kan inte heller själv tänka sig att skriva en dataartikel:

Nej, jag kan tycka att det är bortkastat att endast publicera data utan att analysera och reflektera över resultatet själv. När jag planerar ett forskningsprojekt så finns det i regel en frågeställning i botten, som även kan bygga på tidigare projekt eller kunskap inom området. Att då endast generera data för att inte fullfölja frågeställningen eller sin egen nyfikenhet kan jag inte tänka mig i nuläget i alla fall.⁶⁸

Att det egna lärosätet skulle ha policyer för datadelning är det bara en av de intervjuade som har reflekterat över:

En policy [från forskarens egen sida, författarens anmärkning] har varit att dela data och resultat från skattefinansierade projekt i så stor mån som möjligt. En annan policy rör universitets tre huvuduppgifter: forskning, utbildning och public outreach, där delning av forskningsdata passar in på samtliga uppgifter.⁶⁹

Krav på datadelning i forskningsanslag är heller inget som har påverkat forskarna i någon större utsträckning: ”Nej, jag tror inte det var ett krav. Men själv hade jag med det, för jag tycker att man får ju visa på ... Ja, jag tror det är bra argument för att folk ska ge pengar. Det är inte bara för det här lilla, utan det har en vidare användning också”.⁷⁰

En forskare anser att policyer får mer och mer betydelse, särskilt i USA och Kanada ”där allting som samlas in med skattemedel är offentlig information”.⁷¹ Samma person och ytterligare en forskare anser att kontrollen av efterlevnad är för dålig i Sverige: ”Fler forskningsprojekt förbinder sig [att tillgängliggöra sina data], men jag tror inte kontrollapparaten fungerar. Det borde de se över, de skulle kunna kräva in det i slutredovisningen, ett intyg från SND”.⁷² Att nästan alla vetenskapliga tidskrifter har börjat kräva att data delas ses som positivt, men det är delvis en ”slamkrypare” påpekar en forskare:

[J]ournalerna kräver egentligen inte att man publicerar all rådata – de kräver att man publicerar all information man behöver för att återskapa en graf, vilket ju inte är samma sak. Jag skulle gärna se att framförallt de här High Impact-journalerna skulle kunna ställa krav på att folk faktiskt publicerar all sin data, inte bara det som krävs för att göra grafen.⁷³

5.5 Hur går det till att dela data?

Forskarna har fått kännedom om möjligheten att publicera dataartiklar genom olika kanaler. ”Jag tror att jag såg en notis på någon hemsida. Via Nature – det var där vi publicerade vår artikel – om de gjorde lite reklam för det på sin hemsida eller om det kom upp på Researchgate”, berättar en forskare.⁷⁴ En annan, som skrev sina första dataartiklar redan 2013, hörde talas om möjligheten på en konferens där chefredaktören för den aktuella, då relativt nystartade, datatidskriften deltog: ”[H]an är också chef för FN:s klimatforskningsorgan, och han var ju bra på att få ut informationen, han höll en presentation där han pitchade idén och han nådde en bred publik”.⁷⁵ I ett fall har kollegor tidigare skrivit liknande artiklar⁷⁶ eller så har handledaren kommit med förslaget:

Jag hade lagt ganska mycket tid av mitt avhandlingsarbete på att skapa ett dataset, vilket också innehöll unik och användbar information. Jag och min handledare tyckte därför att ett data paper skulle passa för att beskriva datasetet i detalj.⁷⁷

Några tidskriftsförlag har väldigt strikta mallar för hur en dataartikel ska läggas upp: ”Abstract får vara si och så många ord, introduktionen si och så många ord. Metoden ... där tror jag ... var det ingen begränsning, den får vara hur stor som helst”.⁷⁸ I andra fall har förlaget gett mer generella råd om vilka segment som bör vara med. En forskare berättar att det i början var svårt att förstå hur dataartikeln skulle utformas, men att en kollega hade skrivit en liknande artikel som de kunde hämta inspiration från. Tidskriftens mallar däremot gav inte så mycket ledning:

[N]är man skickar in sin artikel så kan man välja om det ska vara en vanlig forskningsartikel eller om det ska vara ett letter eller andra artikeltyper och där hade de en som kallades för just databas eller ja, *data base*. Och så läste jag en annan artikel från samma tidskrift där de hade beskrivit just en databas – med annan typ av information då, men lite liknande – så då tänkte jag att det där kanske kan passa. Men sen var det helt fritt hur man skulle välja olika rubriker och så. /.../ Bakgrund skulle finnas, men sen var det mera ... ja, texten i mitten, den fick man bestämma själv så det var ganska fritt faktiskt.⁷⁹

En av forskarna tycker att även om fördelarna med dataartiklar överväger så kan det vara en nackdel att ”när man inte involverar dem som har gjort datan så tappar man väldigt mycket av expertisen kring den”.⁸⁰ För att förebygga problem har samma forskare lagt mycket krut på att i den egna dataartikeln diskutera osäkerheten och begränsningarna med den data, kopplad till kartor, som beskrivs:

Användare tenderar ibland att vara lite för okritiska till kvaliteten och göra saker som det kanske inte riktigt är lämpat för. Till exempel det här är ju väldigt grovhuggen data för analyser av hela [geografiskt område], men ibland så gör folk väldigt lokala analyser för någon liten del av något område där vi inte ha den precisionen i våra mätningar så att det är försvarbart.⁸¹

Att dela data är alltså tidskrävande som redan har beskrivits, eftersom det tar tid att strukturera dem och förse dem med metadata. Själva tekniken runt delningen behöver dock inte upplevas som besvärlig, även om ett par av forskarna önskar att universiteten, och i ett fall specifikt biblioteken, kunde hjälpa till mer med digitalisering och stöd för publicering av data: "[D]et är lite svävande nuförtiden vad biblioteken har för roll, och datahantering är ju nånting som jag tycker passar in där".⁸² Andra har fått mycket hjälp med att förbereda sina data för deponering av exempelvis SND eller IT-personal knutna till det egna projektet eller institutionen. En forskare berättar att tidskriftsförlaget krävde att datasetet flyttades till deras server och så att säga låstes under peer review-processen innan det kunde lagras permanent i en portal på forskarens institution. "[E]ftersom vi talar om hundratals gigabyte så tog det ganska lång tid att ladda upp det och döpa filer och man kunde inte ladda upp alla på en gång ... Så den tekniken var ju lite klurig".⁸³ Ett annat önskemål är att det skulle finnas tydligare information om hur det går till att skriva en dataartikel och kanske ett verktyg som förenklar byråkratin runt att dela data: "Men det beror väl på hur data ser ut. Om man har siffror och tabeller som man vill dela med sig av då kanske det är enklare att skicka in till någon grupp eller någon tidskrift som har en arbetsgrupp som sammanställer det där på något sätt och så får man godkänna att det ser bra ut".⁸⁴ En forskare tror att det skulle vara gynnsamt med en stor databas som omfattar flera områden, både nationellt och internationellt "där man kan ta del av öppen data genom en personlig profil (ej anonymt)".⁸⁵

Både vad gäller granskning och citering av data varierar praktikerna något runt vad det egentligen är som granskas och citeras – dataartikeln eller bakomliggande data. Om granskningsprocessen säger en forskare: "Ja, du kan ju antingen vara en väldigt ambitiös granskare som också ber att få datan, så du kan granska den själv. Det har jag nog aldrig gjort. Utan då är det nog snarare så att du tittar på, vad är det de påstår att de här data gör [...]".⁸⁶ Att det egentligen är data som citeras och inte själva dataartikeln råder det större samstämmighet om. "[D]et är väl ett krav så att säga. [...] Det är väl om man tar metoddelen bara som man skulle kunna citera ett data paper utan att använda [bakomliggande data]. Men det har jag aldrig gjort".⁸⁷

Alltså de gångerna jag citerar en dataartikel så är det oftast så att jag har använt datan i analysen, så då har man ju laddat hem datan och använt den som en del i en bredare analys för att få någon slags kontext till sina egna data kanske. Och sen då citerar man dataartikeln. /.../ Jag har citerat den globala kolbudgeten utan att faktiskt använda deras data för de presenterar väldigt mycket siffror i sin artikel. Så det har jag gjort faktiskt.⁸⁸

Dataset kan också publiceras som de är, utan någon dataartikel. En av de forskare som publicerat flera dataartiklar har även varit medförfattare vid ett antal tillfällen när mindre dataset lagts upp i ett europeiskt datarepositorium. "Och om det citeras så – i alla fall i några sökmotorer typ Google Scholar – dyker det upp som en citering. Men det dyker inte upp som en citering i Web of Science till exempel".⁸⁹

5.6 Vad innebär det att dela data?

Att en dataartikel kan granskas, citeras och bli en del av det akademiska meriteringssystemet på samma sätt som en konventionell artikel ser en majoritet av intervjupersonerna som positivt. ”Det beror nog på forskningsfält, men i allmänhet borde det ha stor betydelse för spridning av data. Det går också att se på mängden artiklar och den höga impact-factorn som vissa datatidskrifter har fått, till exempel Scientific Data”.⁹⁰

Att öppna data kan gynna forskningen i stort är ett viktigt motiv för flera forskare att sprida sina data. Det kan vara så att det inte finns ”tid eller pengar att göra forskning på all data” och då är det en fördel om andra kan använda sig av dessa data och eventuellt med ”andra infallsvinklar”.⁹¹ En forskare säger till och med att ”[o]m man ser på det stora hela, om nån råkar stjäla min idé, ja så kanske det ändå tar vetenskapen framåt”.⁹² Samtidigt trodde forskaren att risken att detta skulle hända var liten ”om man har publicerat ett dataset”.⁹³

Ingen av forskarna uttrycker någon rädsla för att utomstående ska missbruka deras öppenhet – i alla fall inte kollegor inom det egna forskningsfältet: ”Det tror jag är en väldigt liten risk i vår disciplin. Det är möjligen att det finns sådana risker i vissa naturvetenskapliga discipliner ...”.⁹⁴ Tvärtom tror forskaren att ju tidigare data tillgängliggörs, desto mer uppmärksamhet och citeringar får upphovspersonen. En annan forskare bekräftar den bilden:

Jag tror att det är väldigt få som upplever att det har varit något problem att folk tar deras data och gör en massa konstiga saker. Oftast, det som händer när folk delar sina data är nog bara är att de får ut mer av det för att någon är intresserad av att använda deras data, men kanske känner att de gärna skulle vilja ha någon med som vet mer om det och då blir de inbjudna att samarbeta.⁹⁵

Den forskare som hade publicerat en dataartikel redan före den publicerade artikeln (se avsnitt 5.3) sade sig ha kontakt med alla forskargrupper i världen som höll på med ämnet och behövde därför inte känna sig orolig:

Det är inte så många fler än att jag visste ungefär att ingen ... Alla visste att jag höll på med det här. Så det var ingen risk att någon skulle ta det, om man säger så. Alla skulle veta också att någon hade tagit min data. Det finns ju andra fält där forskningsfälten är så stora att man inte har överblick [...] och då blir det ju mer riskabelt.⁹⁶

Forskarna i föreliggande studie anser överlag att den peer review-process som en dataartikel genomgår har stor betydelse. ”Ja, för mig så betyder det allt. Jag publicerar inte – annat än på svenska i läroboksformat – sådant som inte är peer reviewed. Så ett data paper som inte är peer reviewed ... Jag kan inte ens komma på några exempel på att jag tittat på det”.⁹⁷ Granskningsprocessen innebär en kvalitetsstämpel eftersom granskade dataset ofta är bättre beskrivna och det går att hitta mer information om dem. Forskaren konstaterar att ”man skriver inte ett data paper om man inte tror att ens egna data är bättre än alla andras. Men hur mycket fog man har för det kan ju variera och där fyller ju peer reviewed en viktig funktion”.⁹⁸ Dessutom ses det som betydelsefullt att någon mer än författarna har gått igenom datasetet och ”sett till att det inte har blivit alltför skevt”.⁹⁹

Jag skulle vara väldigt försiktig med att plocka upp ett dataset ... om det är en grupp jag inte känner till och det inte finns ett peer reviewed data paper som beskriver det – då skulle jag nog vara tveksam till att använda det. Om jag inte själv hade möjlighet att utvärdera kvaliteten på datasetet.¹⁰⁰

Två forskare sade sig också ha fått värdefulla kommentarer från granskarna. Peer review-processen ”förbättrade både datasetet och beskrivningen av datasetet. Totalt lade jag ner cirka fyra veckor för att förbättra datasetet och artikeln efter deras kommentarer”.¹⁰¹ En forskare uttrycker att granskningen bidrog till att slutresultatet blev mer logiskt. ”Om man bara hade lagt in det, då hade det varit vad det var. Nu får man communityn bakom sig så det känns som att det är ett starkare dataset när det är peer reviewed”.¹⁰² Beskrivningar av dataset i en vanlig artikel är också ofta mindre omfattande då det viktiga ”anses vara resultatet av datan, inte datan i sig”.¹⁰³

Citeringar lyfts fram som ett sätt att ge erkännande för det stora arbete det ofta innebär att dela data. ”Det är ju när man själv har varit i branschen och samlat in mycket egna data utan att ... då månar man ju mycket om andra som har gjort det också”.¹⁰⁴ När upphovspersonerna inte har angett hur de vill få erkännande för sina data försöker samma intervjuperson ”göra det på ett sätt som så att säga maximerar den ära man ger dem”.¹⁰⁵ Ibland har dock tidskrifterna synpunkter på citeringsförfarandet och vill inte att skribenterna ska ”citera ett dataset på samma sätt [som en vanlig artikel]”.¹⁰⁶ Det finns en rörelse mot att försöka etablera konsekventa former för datacitering inom vissa forskningsfält, men ännu finns inga självklara sätt att göra det på:

Det kan stå i en fotnot i artikeln utan att det är en referens i artikeln som sedan länkas till Web of science och Google Scholar och då får man ju inga citeringar på det. Det kan vara så att de använder något annat än det man har bett dem om – de kanske har sett våra data i en bok och så citerar de den boken, men inte våra data.¹⁰⁷

En dataartikel kan, enligt en intervjuperson, underlätta citeringsförfarandet ”för att man är van vid att citera artiklar eller papers och det går snabbare eftersom man vet hur man ska göra det”.¹⁰⁸ Men inte heller runt dataartiklar är citeringsformerna helt stabila: ”[V]issa vill att man ska citera ett data paper, andra citerar data och inte data paperet och vissa gör både och”.¹⁰⁹ En forskare tycker att det för egen del känns som att en dataartikel medför ett större erkännande för det arbete som lagts ner, på grund av att ”det blir mer citerbart och det ser bättre ut i ett cv att ha en artikel och inte bara ett dataset”.¹¹⁰ Dataartiklar gör det också lättare att hitta beskrivna data eftersom de ofta är ”bättre indexerade”¹¹¹ och därmed kommer ”högre upp i sökmotorernas algoritmer”¹¹²: ”De dataset som har ett data paper kommer att få mycket högre sök-cred av Google Scholar till exempel”.¹¹³

Att det finns vissa dataartiklar som är väldigt väl citerade kommenterar en intervjuperson så här: ”[D]et är för att man kopplar ihop de där två sakerna; när du använder mina data så citera mitt data paper. Och om det då är data som inte bara är bra, utan dessutom gör något som ingen annan har gjort tidigare, då kan det bli väldigt mycket citeringar”.¹¹⁴ En väl citerad dataartikel kan förbättra karriärmöjligheterna på flera sätt; den innebär ett erkännande av det arbete som lagts ner, kan läggas till på ett cv och kan användas i framtida medelsansökningar.

Den [dataartikeln] är väldigt väl citerad och det har gjort att jag har haft större chans att få mer pengar. Och sen får man ganska mycket cred i forskarvärlden om folk vet att man ligger bakom ett dataset som

faktiskt är bra och som folk använder. Folk uppskattar det. Så det har varit bra för min fortsatta karriär.¹¹⁵

En annan forskare säger att dataartikeln ”blir en bekräftelse på att det är lika värdefullt som en annan vetenskaplig artikel /.../ och hög citeringsfrekvens brukar ju anses av vissa att det speglar publikationens vetenskapliga kvalitet så på det sättet skulle mycket citeringar kunna vara positivt”.¹¹⁶ Samtidigt är det med dataartiklar precis som med konventionella artiklar – några få blir flitigt citerade medan andra bara citeras av några få eller i värsta fall inte alls. Så ”det är inte säkert att det är ett bra sätt att bli väldigt citerad på”.¹¹⁷ En intervjuperson konstaterar om sitt data paper att det ”i nuläget endast [är] citerat av artiklar som jag själv har skrivit”.¹¹⁸ Det hindrar inte att svaret på frågan om forskaren skulle kunna tänka sig att skriva fler dataartiklar blir ”Ja, definitivt”.¹¹⁹ En annan forskare har haft nytta av att kunna citera den egna dataartikeln vid föredrag på konferenser och vid medelsansökningar: ”Det blir väl lite mer vetenskaplig höjd på det eller hur ska man uttrycka det... I och med att det finns publicerat så kan man ju också citera och använda den [dataartikeln] ... Ja, när man ska söka medel eller liknande, använda den som en referens, visa att det här har vi gjort”.¹²⁰

6 Viktiga frågor vid delning och publicering av data

Syftet med studien var att skapa fördjupad förståelse för dataartiklars betydelse i forskningsprocessen och på så sätt bidra till kunskapen om öppna forskningsdata som en del av vetenskaplig kommunikation. Åtta forskare utfrågades om sina erfarenheter och i föregående kapitel presenterades resultatet uppdelat i sex teman som identifierades under en initial kodning av materialet. I detta kapitel analyseras det empiriska materialet med utgångspunkt i det teoretiska ramverket och sätts även i relation till tidigare forskning. Genom att utgå från det sociotekniska perspektivet och betrakta dataartikeln som en stabiliserande socioteknisk ensemble sätts fokus på processer som sker i samspel mellan människor, teknik och praktiker: Nya tekniker för att samla in, lagra och dela data ställer nya krav på inte bara tekniska artefakter och infrastrukturer för vetenskaplig kommunikation utan även på de sociala aktörerna. Dataartikeln har uppstått som ett sätt – av flera – att möta dessa krav och det är intressant att undersöka om och i så fall hur den kan bidra till att stabilisera praktikerna runt datadelning. Hur en socioteknisk ensemble tolkas, utformas och används är ingen självklarhet utan något som avgörs i en förhandlingsprocess där behoven hos olika grupper av användare och tekniska lösningar har betydelse (Francke 2008). Särskilt fokus läggs därför i analysen på de sociala praktikerna runt dataartikeln, samspelet mellan människor, teknik och artefakter samt dokumentets funktioner. Intervjuresultaten sätts även i relation till relevant forskning. Kapitlet är indelat i tre övergripande teman som framträdde när de transkriberade intervjuerna kodades i en andra bearbetning av materialet. Dessa tre teman är *”Förtroende och trovärdighet”*, *”Samarbete och erkännande”* samt *”Påverkan och status”*.

6.1 Förtroende och trovärdighet

Ett sociotekniskt perspektiv kan användas för att belysa hur den tekniska utvecklingen kontinuerligt förändrar förutsättningarna för socialt samspel och hur tekniken i sin tur utvecklas till följd av användarnas behov av bland annat nya sätt att kunna samarbeta. Förtroende och trovärdighet berör två – delvis sammanflätade – processer som alltid har varit viktiga när forskare har utbytt data och som både har påverkat och påverkats av den tekniska utvecklingen. Den första processen rör framför allt hur förtroende skapas forskarkollegor sinsemellan, men även informanternas förtroende för forskare. Den andra gäller hur värdet och kvaliteten hos egna och andras data kan valideras – vilken trovärdighet de har. Båda processerna framstår som centrala för de intervjuade forskarna vid delning och publicering av egna data och återanvändning av andras. Tidigare forskning bekräftar att frågorna kan ha betydelse för attityder och handlande när det gäller öppna data (Hickson et al. 2016; Park & Wolfram 2017; Tenopir et al. 2011): Till exempel kan rädsla för att andra forskare ska ”stjäla” forskningsidéer eller på andra sätt missbruka öppenheten vara ett hinder för datadelning. Därför var det intressant att en av de intervjuade forskarna hade valt att publicera en dataartikel redan innan den applicerade artikeln

var färdig. En av förklaringarna till detta var att forskningsfältet inte var större än att "alla skulle veta" om någon tog denna data.¹²¹ Även i flera andra intervjuer återkommer utsagor om att risk för missbruk av data är något som möjligen kan finnas inom andra discipliner och forskningsområden. Intervjusvaren tyder på forskarna upplever att det finns en kultur inom det egna fältet som möjliggör öppenhet: Att dela och publicera data tidigt ses inte som en risk utan i första hand som något positivt som kan ge större uppmärksamhet, fler citeringar och öppna nya möjligheter till samarbete. Forskarna känner alltså förtroende för sina kollegor inom fältet, trots att det kan röra sig om forskargrupper som är utspridda i stora delar av världen och där det är omöjligt att ha personlig kontakt med alla. Det är möjligt att dataartikeln, genom att den i likhet med en konventionell artikel bidrar till att ge en översiktlig bild av kunskapsläget inom ett forskningsfält och placerar in data i relation till andra, bidrar till att skapa denna känsla av social kontroll. En annan tolkning är att digitaliseringen inte bara har möjliggjort nya sätt att samarbeta utan också har bidragit till att nya typer av *confidence pathways* (Knorr Cetina 1999, se avsnitt 3.1) växer fram. Dessa är inte, som i Karin Knorr Cetinas (ibid.) beskrivningar, beroende av att forskarna har arbetat i samma laboratoriemiljö, deltagit i samma experiment och därmed delar historia. Visserligen är det fullt möjligt att två forskare kan bidra med data till samma dataartikel utan att ha någon kontakt sinsemellan och på så sätt vara del av ett samarbete i vad som kanske skulle kunna beskrivas som en gemensam virtuell laboratoriemiljö. Men förtroendet sträcker sig längre än så eftersom det tycks omfatta även kollegor som intervjupersonerna aldrig har samarbetat med. Kanske är det i spåren av digitaliseringen och kulturen runt öppna data rentav möjligt att tala om framväxten av nya epistemiska kulturer som förenas inte så mycket av att deltagarna *har samarbetat* och delat praktiker i sitt dagliga arbete utan av att de delar praktiker runt delning av data och därmed kanske en dag *kommer att samarbeta*?

Att informanter och personer som olika sätt förekommer i forskningsmaterial ska kunna känna förtroende för forskare är viktigt inte bara i det enskilda projektet utan även på ett större plan för att möjliggöra fortsatt forskning. När data delas öppet måste forskaren inte bara fundera över hur materialet hanteras i det aktuella forskningsprojektet utan även ta ansvar för hur materialet kan komma att användas i framtiden. Detta har inte minst visats av att ny teknik bland annat kan göra det möjligt att återidentifiera uppgifter som tidigare betraktats som avidentifierade (se avsnitt 2.2.2). Frågor om etik och integritetsskydd är centrala för de intervjuade forskarna i avgörandet av vilka data som delas och inte delas. Men känsliga uppgifter behöver inte per automatik innebära att material inte kan delas. I vissa fall kan sådana uppgifter som går att koppla till individdata tas bort. Två av de forskare som deponerat data i SND:s databas har använt sig av, eller övervägt att använda sig av, möjligheten att tillgängliggöra data med restriktioner (till exempel under förutsättning att författarna själva får godkänna framtida användning av materialet).¹²² Den forskare som hade tillgängliggjort alla sina data utom känsligt material om lokalpolitiker hade till exempel funderat på sätt att kunna deponera även detta: "Jag tror att man kan belägga det med restriktion i SND. Eller så får jag rensa ut vissa uppgifter".¹²³ Här har alltså människors behov drivit fram tekniska lösningar för att "spärra" användning av data tills upphovspersonen gett sitt godkännande. Ingen av de forskare som skrivit dataartiklar nämnde att de skulle ha deponerat sina data med särskilda villkor för framtida användning. En möjlig förklaring är att de data som beskrivs i dataartiklarna oftare är kvantitativt material som inte går att koppla till enskilda individer, medan de forskare som deponerat data hos SND i större utsträckning har delat kvalitativa data där det kan finnas innehåll som är känsligt ur integritetssynpunkt. Men författaren till en dataartikel kan också i texten "tala

med” kollegor som vill använda beskrivna data och på så sätt ge förslag på framtida forskning, förklara begränsningar och även avråda från olämpliga sätt att använda materialet. Dataartikeln blir just ett sådant dokument som kan nå ut utanför laboratoriet och bli en resurs i andra forskningskontexter, som Bernd Frohmann (2004) beskriver. Kanske kan dataartikeln i någon mån ersätta både det sociala samspelet i personliga kontakter och de tekniska möjligheter SND ger för att i databasen markera olika villkor för datadelning. Genom att citeringsförfarandet formaliseras blir det också enklare att följa och kontrollera hur delade data används av andra, även om kontrollen förstås då inte kan ske förrän i efterhand.

Förtroende – eller snarare brist på förtroende – för det kunskapsmässiga värdet och kvaliteten hos både egna och andras data är en annan faktor som kan förhindra delning och återanvändning av data. Tvivel på att egna data har något värde för andra är till exempel en återkommande förklaring till att data inte delas (Tam 2016). Detta tror också flera av de intervjuade forskarna i denna studie är en vanlig orsak till varför forskare inte delar data. Själva går de delvis olika vägar: Några funderar inte så mycket över frågan utan säger sig dela alla data av princip – så länge det inte finns etiska hinder. Visst kan det vara så att deras data inte har något värde för andra ”men det vet man inte” [i förväg].¹²⁴ Andra delar inte alla data utan bara dem som de tror kan vara användbara för andra. Att dataartiklar går igenom en peer review-process har stor betydelse: flera av de juniora forskarna beskriver hur granskarnas synpunkter förbättrat deras dataset.¹²⁵ I några fall gjordes bara mindre ändringar, men en forskare lade ner ”cirka fyra veckor för att förbättra datasetet och artikeln” efter granskarnas kommentarer. Det beskrivs också som en styrka att ha ”communityn bakom sig”.¹²⁶ En av de seniora forskarna säger att peer review-processen betyder allt och skulle aldrig själv publicera sig – annat än i läroboksformat – utan att texten har gått igenom en sådan process. Samma person kan inte komma på något exempel på att ens ha tittat på en dataartikel som inte har genomgått peer review.¹²⁷ Flera av intervjupersonerna skulle vara tveksamma till att citera ett dataset eller en dataartikel som inte granskats – om inte en väldigt välkänd forskargrupp eller möjligen en myndighet stod bakom materialet. Det finns alltså en skillnad mellan delad, publicerad och Publicerad data (Callaghan et al. 2012; Lawrence et al. 2011) där Publicering i form av en dataartikel inte bara ger data ökad trovärdighet utan också kan förbättra dess kvalitet och därmed även öka värdet för framtida användare. På samma sätt som vi såg ovan är det möjligt att en dataartikel skulle kunna ersätta personliga kontakter eftersom peer review-processen ökar möjligheten att känna förtroende för innehållet oavsett om de som vill återanvända data känner (till) upphovspersonerna eller inte. Dataartikeln skulle därmed kunna påverka de sociala praktikerna runt hur en forskare bedömer vilka kollegors data som går att lita på.

6.2 Samarbete och erkännande

Att dela data på ett sätt som gör dem användbara för andra kan vara mycket tidskrävande utan att nödvändigtvis ge någon direkt lön för mödan. Enligt Chavan och Penev (2011) har en dataartikel tre huvudsyften: att bidra till att forskaren får erkännande för sitt arbete med att dela data, att beskriva data och att sprida kännedom om data i forskarsamfundet. Det första syftet handlar alltså om *erkännande*, medan de båda andra kan sägas utgöra förutsättningar för att *samarbete* runt delade data ska kunna uppstå. Intervjuerna visar att båda faktorerna kan påverka och i sin tur påverkas av datadelning och publicering av dataartiklar. Dokumentperspektivet (se avsnitt 3.3) gör det möjligt att sätta fokus på materi-

ella aspekter av en dataartikel samt på de aktiviteter som sker runt omkring den: I vilken kontext skapas och används dataartikeln, på vilka sätt och av vem? Det är även intressant att diskutera dokumentets stabiliserande roll i det att artikeln kan placera in en forskare i förhållande till andra på den akademiska kartan (Frohmann 2004; Leonelli 2016) vara en del av det akademiska belöningssystemet (Frohmann 2004), och ett sätt att stabilisera processerna runt datainsamling och analys genom att passa in dem i en sammanhängande berättelse (Leonelli 2016).

Tack vare den tekniska utvecklingen skapas hela tiden nya möjligheter till samarbete och de kollegor en forskare samarbetar tättast med kan idag finnas nästan varsomhelst i världen. Genom dessa samarbeten kan också nya praktiker runt delning av data utvecklas. I flera intervjuer framkommer att framgångsrika forskare kan välja vilka kollegor de vill samarbeta med och ett sätt att få dåligt rykte i detta sammanhang är att vara snål med data. Forskare som redan är väletablerade och vet att de får fortsatt finansiering behöver inte bry sig så mycket om detta, men en återkommande bild är att det är omodernt att hålla på sina data och att det är ett arbetssätt som inte fungerar för den som vill göra akademisk karriär idag. Detta bekräftar Knorr Cetinas (1999) beskrivning av hur forskare delar in sina medarbetare i grupper, inte bara utifrån trovärdighet och vems resultat som går att bygga vidare på, utan även utifrån vilka kollegor forskaren vill samarbeta med eller undvika.

Att ge erkännande för arbetet det innebär att samla in och tillgängliggöra data är en viktig fråga för intervjupersonerna: Den som själv delar data är också mån om att andras arbete ska erkännas. Här handlar det inte bara om utomstående forskare utan även om att juniora forskare i den egna forskargruppen – som ofta är de som samlar in rådatan – själva ska hinna publicera artiklar baserade på dessa data innan de delas fritt. Att en doktorand uppmanas att skriva en dataartikel kan då innebära att användbara data snabbare görs fritt tillgängliga för alla samtidigt som doktorandens arbete med datainsamling får ett formellt erkännande. När det gäller det erkännande forskaren själv får finns det en viktig aspekt med dataartiklar utöver den som gäller det akademiska meriteringssystemet med citeringar och artiklar på cv:t som kan verka positivt för fortsatta karriär- och finansieringsmöjligheter. Denna aspekt rör det mer informella erkännande en forskare kan få från kollegor inom fältet. Några av de intervjuade forskarna har, själva eller i samarbete med kollegor, tagit på sig arbetet med att sammanställa data från många källor i en stor databas. Dessa databaser har sedan beskrivits i dataartiklar där alla som bidragit med data också har erbjudits att stå som medförfattare. När det gäller författarskap och citeringar får alltså den eller de forskare som har gjort det största arbetet inte större erkännande än forskare som kanske bara bidragit med ”en enda mätpunkt av femhundra”.¹²⁸ Däremot är det troligt att det stora arbetet ger erkännande bland de forskargrupper som arbetar med samma ämne, något som i förlängningen också kan vara gynnsamt för karriären, bland annat genom att det skapar nya möjligheter till samarbete.

Även data som publiceras utan att ha beskrivits i en dataartikel kan ge författaren många citeringar, men flera forskare beskriver hur formerna för detta varierar kraftigt. Att delade data används utan någon form av erkännande alls förefaller enligt intervjuerna vara ovanligt. Det förekommer dock att upphovspersonerna anges i en fotnot i stället för i referenslistan och därmed syns inte hänvisningen som en citering i exempelvis Web of Science eller Google Scholar. När det finns många upphovspersoner till ett dataset är det inte heller säkert att alla kommer med i referenslistan – antingen för att författaren inte skriver ut

alla namn eller för att förlagen stryker namn av utrymmesskäl. Detta kan särskilt missgynna författare vars efternamn har en begynnelsebokstav som kommer sent i alfabetet och juniora forskare som – i vissa discipliner – ofta står längre ner i författarlistan. Erkännande är alltså viktigt, men av intervjuerna framkommer också att inte alla krav på erkännande ses som rimliga. Framförallt gäller detta kollegor som kräver att få stå som medförfattare för att dela data eller ”ser till att det endast är de som förstår hur datasetet är strukturerat, vilket också resulterar i medförfattarskap i de artiklar som vill använda datasetet”.¹²⁹ En intervjuperson berättar att det inom dennes disciplin av många till och med ses som oetiskt att stå som medförfattare utan att ”faktiskt ha varit med och skrivit på artikeln eller gjort analyser som ligger till grund för artikeln”.¹³⁰ Visserligen finns en förståelse för kravet på erkännande eftersom det kan ligga flera års arbete bakom vissa dataset, men en dataartikel som ger citeringar ses som en mer rimlig form än ett medförfattarskap. Detta är ännu ett exempel på hur en dataartikel skulle kunna bidra till att stabilisera processerna runt delning av data.

Enligt det sociotekniska perspektiv som studien anlägger skapas en artefakts betydelse och funktion i samspel mellan sociala praktiker och tekniska förutsättningar. Därmed kan artefakter också få olika betydelser för olika grupper. Ett exempel på detta som framkom i intervjuerna är hur synen på dataartikeln skiljer sig åt mellan de intervjupersoner som själva hade skrivit sådana och den forskare som enbart hade citerat en dataartikel. Den sistnämnda ansåg att det var ett ”lättvindigt sätt att addera saker på den akademiska meritlistan” och skulle tycka att det var ”bortkastat att endast publicera data utan att analysera och reflektera över resultatet själv”.¹³¹ En annan forskare hade tvärtom valt dataartikeln som publiceringsform just för att det var positivt att kunna få ut datasetet snabbt, utan att behöva lägga så mycket tid på att skriva en diskussion.

Till skillnad från andra sätt att dela data ger dataartikeln erkännande till upphovspersonen oavsett om delade data används och citeras eller ej, eftersom den kan sättas upp på forskarens cv tillsammans med konventionella artiklar och även räknas i bibliometriska eller andra utvärderingar av genomslag och produktivitet. Även den forskare, vars dataartikel när intervjun gjordes inte hade citerats av någon annan än forskaren själv, tyckte att arbetet med dataartikeln hade gett lön för mödan och skulle ”definitivt”¹³² kunna tänka sig att skriva fler. Detta bekräftar dataartikelns betydelse som en del av det akademiska belöningsystemet (Frohmman 2004).

6.3 Påverkan och status

Sociala praktiker uppstår runt människors dagliga aktiviteter och interaktion med varandra och samspel med material och andra resurser (Van House 2003b). Genom att använda en praktikorienterad ansats och begreppet epistemiska kulturer kan undersökningen av forskares användning av dataartiklar sättas i ett sammanhang där individens samspel med omvärlden och med olika tekniska verktyg är centrala och ser olika ut i olika sammanhang. I några intervjuer beskrivs kulturen runt datadelning som något som forskaren socialiserats in i som doktorand, och som kopplad till framförallt handledaren, men även till ett visst forskningsprojekt.¹³³ Men en intervjuperson kunde inte minnas att frågor runt öppna data någonsin tagits upp bland kollegorna på institutionen. I stället var det kulturen internationellt, bland kollegor inom det egna forskningsämnet, som gjort det till en självklarhet att dela data. Forskaren påpekade också att ämnets särart (ett smalt ämne som det finns väldigt lite material om) hade bidragit till en vilja att dela data.¹³⁴

Detta visar hur samspelet med andra – i små eller stora forskargrupper – har central betydelse för de sociala praktiker som växer fram. Att även beskaffenheten hos de data som samlas in har stor betydelse för vilka sociala praktiker som utformas runt datadelning, visade Winnie Tams (2016) undersökning av två underdiscipliner inom geografi (se avsnitt [2.1.3](#)). Detta kan förklara varför intervjuresultatet visar på att aktiviteterna runt datadelning kan skilja sig kraftigt åt även på institutionsnivå och att det främst är runt själva forskningsprojekten som enhetliga kulturer runt datadelning uppstår. Det kan också röra sig om en kultur bland yngre forskare i början av sin akademiska karriär. Eller bland alla forskare som anser det vara självklart att dela data. Inte i något fall har en intervjuperson pratat om att det skulle finnas gemensamma praktiker runt öppna data på institutionen och än mindre på lärosätet. När det talas om enhetliga kulturer på institutionsnivå gäller det i stället exempelvis hur forskaren väljer – eller styrs mot – att publicera sig i vissa format (artiklar som är peer reviewed dominerar¹³⁵). Uppslaget att publicera en dataartikel har i något fall kommit ”uppifrån” från en doktorands handledare, men annars har idén kommit från kollegor, från en föreläsare på en konferens eller i ett fall en annons.¹³⁶

Att inte dela med sig av sina data beskrivs som ett omodernt arbetssätt som bara de kan tillåta sig som har jobbat länge inom ett fält och ”vet att de är tillräckligt kända och väletablerade så att de får fortsatt finansiering ändå”.¹³⁷ Det råder samstämmighet om att en sådan attityd inte fungerar för den som vill göra akademisk karriär idag. Detta skulle kunna bekräfta fynden i Van den Eyndens och Bishops (2014) undersökning som visade på att forskare har olika skäl för att dela data i olika skeden av sin karriär och att utbildning om datadelning i början av forskarkarriären kan ha en stor roll. Forskare som själva är beroende av att andra delar data anser också att det är viktigt att föregå med gott exempel. Att de intervjuade forskarnas praktiker runt delning av data hittills verkar ha formats utifrån smala forskningsområden snarare än top-down på institutioner eller lärosäten är kanske inte så konstigt med tanke på att policyer runt öppna data har börjat antas först de senaste åren (se avsnitt [2.2.1](#)). En intervjuperson beskriver hur denna utveckling kommit mycket längre i exempelvis USA med striktare uppföljning av att data verkligen delas som utlovat. Av intervjuerna framkommer också att det (ännu) inte är vanligt att forskarna har delat data på grund av att det har varit ett krav från finansiärerna. Däremot tror flera att delning av data har påverkat eller kan komma att påverka deras chanser att få forskningsanslag i positiv riktning, genom att de i ansökan visar att data kan ha en vidare användning utanför ett avgränsat forskningsprojekt. I flera intervjuer uttrycks också åsikten att data inte är personlig egendom och särskilt inte data som samlats in i offentligt finansierade projekt.

Delning av data är i sig inget nytt fenomen, vilket också framkom tydligt i intervjuerna med de två forskare som har varit verksamma längst. Däremot utvecklas hela tiden mer och mer avancerad teknik och metoder för att samla in, bearbeta och dela data i allt större skala och för att formalisera samarbetet – något som i sin tur skapar nya sätt att samarbeta och behov av fortsatt teknikutveckling (Borgman 2015; Van House 2003a). Ett tydligt exempel på detta gav den forskare som beskrev hur utvecklingen gått från att be kollegor skicka in data utifrån en förbestämd mall till att låta en programmerare skapa script²⁰ som automatiskt omvandlar inskickade data till rätt format. Enda förutsättningen var att det måste vara tillräckligt mycket data för att det ska löna sig att lägga tid på att skapa scrip-

²⁰ Ett script är ett enkelt datorprogram som automatiserar arbetsuppgifter som annars skulle utföras steg för steg av användaren. Källa: IT-ord, Computer Sweden.

tet. Forskaren tror att det är ”vägen framåt för att samla in mer data”¹³⁸ och säger att hittills har försöket varit lyckat.

6.4 Sammanfattning

Trots att intervjupersonerna kommer från vitt skilda forskningsområden finns det mer som förenar än som skiljer dem åt när det gäller synen på dataartikelns betydelse och delning av data i allmänhet. Av utsagorna i denna intervjustudie har framkommit att det finns många instabila processer runt delning och publicering av data: praxis saknas inom många forskningsfält beträffande var delade data deponeras; hur de kan göras sökbara; hur de kan granskas, publiceras och citeras; och hur upphovsmännen ges erkännande för sitt arbete. Dataartikeln kan bidra till att stabilisera flera av dessa processer, till exempel genom att göra data lättare att hitta och ge data en kvalitetsstämpel. Genom att formerna för erkännande från såväl kollegor som arbetsgivare och forskningsfinansiärer formaliseras kan dataartikeln också bidra till att ge forskare ett nytt incitament att dela data. Intervjusvaren visar vidare på att kulturen en forskare socialiseras in i verkar ha haft stor betydelse för praktikerna runt delning av data. Intressant är här att en återkommande bild av de sociala praktikerna runt delning av data är att de verkar formas med utgångspunkt i smala forskningsämnen och enskilda projekt snarare än i hela discipliner eller underdiscipliner. Detta bekräftar Van den Eyndens och Bishops (2014) undersökning om attityder till datadelning som visade på stora variationer inom discipliner och även mönster som sträckte sig över disciplingränserna (se avsnitt [2.2.1](#)). Praktikerna kan också se olika ut beroende på hur långt forskaren har kommit i sin karriär. Sammantaget talar intervjusvaren för att forskares användning av dataartikeln skulle kunna ses som ett sådant nätverk av ömsesidigt stabiliserande objekt och aktiviteter som Bijker (1995) kallar socioteknisk ensemble. Detta eftersom dataartikeln kan bidra till att stabilisera såväl praktiker runt delning och publicering av data som förtroendet mellan sociala aktörer (framförallt de forskare som delar och återanvänder data) och trovärdigheten hos tekniska artefakter (de data som delas). Enligt Bijkers (2001, s. 27) tankar om *interpretative flexibility* sker stabiliseringen av en artefakt genom en social process och är därmed föremål för olika val, intressen och värderingar. Om dataartikelns betydelse som dokumenttyp i framtiden kommer att fortsätta öka eller istället minska avgörs enligt detta synsätt varken av enskilda tekniska eller sociala faktorer utan av förhandlingar mellan olika relevanta användargrupper om hur denna artefakt ska användas (Bijker 1995; Francke 2008).

7 Diskussion och slutsatser

Att dela data är långt ifrån någon ny företeelse i forskarsammanhang, något som togs upp redan i inledningens beskrivning av astronomiska upptäckter, och som även har bekräftats av intervjuaren i denna studie. Däremot har digitaliseringen medfört en snabb utveckling av nya metoder att samla in, lagra, dela och återanvända allt större mängder data och dessutom finns det från forskningspolitiskt håll en stark rörelse för att så mycket data som möjligt ska göras öppet tillgängliga. Syftet med denna studie var att skapa fördjupad förståelse för dataartiklars betydelse i forskningsprocessen och på så sätt bidra till kunskapen om forskningsdata som en del av vetenskaplig kommunikation. Eftersom arbetet med dataartiklar är nära sammanflätat med frågor som rör hantering och delning av data har även dessa ämnen berörts i relativt stor utsträckning. Första delen av detta kapitel ägnas åt att redovisa hur forskningsfrågorna har besvarats. Uppsatsen avslutas sedan med en diskussion om undersökningens betydelse och begränsningar samt förslag på fortsatta forskningsfrågor.

7.1 Återkoppling till forskningsfrågorna

Fyra forskningsfrågor ställdes i inledningen av uppsatsarbetet och eftersom svaren delvis glider in i varandra besvaras de här i en sammanhängande text. Frågeställningarna var:

1. Vilka funktioner kan en dataartikel fylla för forskare i deras arbete?
2. Vad skiljer en dataartikel från andra sätt att dela, publicera och återanvända data?
3. Vilken roll kan dataartiklar ha för att möjliggöra och underlätta återanvändning av data?
4. Vilken betydelse har det att en dataartikel kan granskas, citeras och vara en del av det akademiska meriteringssystemet på samma sätt som en konventionell vetenskaplig artikel?

Studien har visat att en dataartikel kan fylla flera olika funktioner. Till de mest centrala hör att synliggöra inte bara data och metoder som forskaren tror kan vara användbara för andra utan även det arbete som forskaren lägger ner på att tillgängliggöra dessa data. Detta bekräftar att dataartikelns tre huvudsyften är att ge forskaren erkännande för sitt arbete, att beskriva data och att sprida kännedom om data (Chavan & Penev 2011). I en dataartikel lägger forskaren tid på att ”strukturera och beskriva data så att de kan förstås utan [...] hjälp”.¹³⁹ Granskningsprocessen en dataartikel går igenom ger också de data den beskriver en kvalitetsstämpel. Därmed kan den bli just ett sådant dokument som Bernd Frohmann (2004) talar om, som kan nå utanför laboratoriet och bli en resurs i andra forskningskontexter. Detta kan också innebära att Frohmans (2004) ifrågasättande

av den vetenskapliga artikeln som bärare av den information som behövs för att producera ny forskning delvis skulle behöva omvärderas. Kanske kan en dataartikel samtidigt överföra både sådan information som placerar in forskaren i förhållande till kollegor och forskningsområdet och sådan som kan användas för att producera ny forskning? Jag skriver *delvis* omvärderas eftersom dataartikelns utformning och den information den kan överföra trots allt ansluter mer till Frohmans beskrivning av praktikdiskurser, där fokus läggs på aktiviteter och materiella aspekter, än till den epistemiska diskursens syn på kunskap som huvudsakligen teoretisk och abstrakt. Samtidigt behåller dataartikeln funktionen hos en konventionell vetenskaplig artikel att ge en översiktlig bild av ett forskningsfält, bekräfta kunskapsläget för den som vill placera ett vetenskapligt arbete i relation till ett annat och som en del av det akademiska belöningssystemet (ibid. 2004).

En dataartikel ska inte innehålla forskningsfynd, analys eller argumentation och kan därför vara ett snabbare alternativ än en konventionell artikel för forskare som i första hand vill ”få ut”¹⁴⁰ data. Därför har dataartikeln i något fall till och med publicerats före den applicerade artikeln, även om parallellpublicering är vanligare. En fördel när dataartikeln publiceras före eller parallellt med den applicerade artikeln är att metoddelen i den applicerade artikeln inte behöver tyngas med alla detaljer runt insamling av data och vald metod. Ofta är textlängden i en artikel begränsad och forskarna ser det då som positivt att i stället kunna ägna större utrymme åt att beskriva själva forskningsfynden. Strategin med parallellpublicering innebär dock en viss risktagning i de fall dataartikeln inte blir antagen i tid eftersom den applicerade artikelns metoddelen i sådana fall måste skrivas om. Överlag kan det vara svårt att få en dataartikel antagen i välrenommerade datatidskrifter, och inom vissa ämnesområden saknas dedikerade datatidskrifter vilket gör det ännu svårare. En annan fördel med att en dataartikel inte ska innehålla några forskningsfynd är att det kan underlätta publicering av negativa eller neutrala resultat: datainsamlings- och analysprocesserna måste inte passas in i en sammanhängande berättelse där de aspekter av forskningsresultaten som inte understödjer argumenten utelämnas (Leonelli 2016). Som exempel nämndes att det vid förskrivning av medicin kan vara viktigt att känna till att vetenskapliga studier har gjorts indelade efter kön även i de fall inga könsskillnader hittats.

Eftersom dataartiklar synliggör både delade data i sig och det arbete forskaren har lagt ner på att ta fram dessa data kan de möjliggöra återanvändning av data på flera sätt. Till att börja med skulle synliggörandet – och i förlängningen erkännandet – av forskarens arbete kunna bidra till ökad motivation att dela data. Här har dataartikelns likhet med en konventionell vetenskaplig artikel stor betydelse, på så sätt att den kan granskas, citeras och vara en del av det vetenskapliga meriteringssystemet. I stället för att forskaren belönas först när data används av någon annan – genom en citering eller medförfattarskap – kan en dataartikel sägas ge direkt belöning då författaren får en ny artikel att sätta upp på sitt cv. Denna belöning utfaller dessutom oavsett om delade data används eller ej. Synliggörandet innebär också att delade data lättare kan hittas även utanför den närmaste gruppen av kollegor på institutionen och inom det egna forskningsområdet, något som kan leda till nya – kanske tvärvetenskapliga – samarbeten eller ny sätt att använda delade data. Detta är särskilt viktigt när det inte finns väl etablerade repositorer som en stor andel av forskarna inom forskningsområdet använder sig av och bevakar. Det formaliserade citeringsförfarandet kan bidra till att stabilisera praktikerna så att erkännande ges på ett sätt som intervjupersonerna uppfattar som ”fair”: inte för lite (som i en fotnot) eller för mycket (som i ett medförfattarskap). Dessutom ger dataartikeln ett incitament för fors-

kare att lägga ner tid på att göra data användbara för andra eftersom de får något igen för detta arbete – oavsett om data används av andra eller ej.

För många av de intervjuade forskarna är det alltså positivt att dataartikeln kan bidra till meritering och kanske generera nya anslag och det finns också rent praktiska fördelar med dataartikeln som att den kan frigöra utrymme i applicerade artiklar. Men mitt intryck är ändå att andra effekter av dataartikeln är minst lika viktiga – till exempel att granskingsprocessen ger delade data en kvalitetsstämpel och att godtyckligheten runt delning av data minskar på så sätt att en forskare inte behöver känna (till) en kollega för att kunna få tillgång till data. När data delas informellt krävs personlig kontakt och i samband med detta kan frågor om tänkt användning, forskningsetik och eventuella begränsningar hos det insamlade materialet diskuteras direkt. Att dela data som är fritt tillgängliga för alla kan antas kräva större förtroende för forskarkollegor och andra potentiella användare eftersom det är svårare att på förhand ha kontroll över och kunna förutse hur data kan komma att användas. Peer review-processen, som en dataartikel går igenom före publicering, innebär att delade data får en kvalitetsstämpel och en doktorand upplevde det som en trygghet att ha forskarsamfundet bakom sig i och med att data granskats före delning. Granskarnas kommentarer har också lett till förbättringar av såväl delade dataset som beskrivningen av dessa dataset. Att delade data är granskade har också stor betydelse för intervjupersonerna när de själva ska använda andras data. Flera säger att de skulle vara mycket tveksamma till att använda delade data som inte granskats – om de inte kom från en välkänd forskargrupp eller möjligen från en myndighet. Detta är ytterligare exempel på hur dataartikeln kan bidra till systemet stabiliseras och att nya sociala praktiker runt datadelning formas.

Enligt Bijkers (2001, s. 27) tankar om *interpretative flexibility* sker stabiliseringen av en artefakt genom en social process och är därmed föremål för olika val, intressen och värderingar. Om dataartikeln betydelse som dokumenttyp i framtiden kommer att fortsätta öka eller istället minska avgörs enligt detta synsätt varken av enskilda tekniska eller sociala faktorer utan av förhandlingar mellan olika relevanta användargrupper om hur denna artefakt ska användas (Bijker 1995; Francke 2008). Här kan även teknikpolitik och policyer få betydelse. Samtidigt är dataartikeln och datatidskriften bara en liten del av ett större sociotekniskt system som inbegriper till exempel de vetenskapliga tidskrifternas och förlagens roll inom vetenskaplig kommunikation, fortsatt teknisk utveckling och policyer för öppen vetenskap. Till exempel införs nu policyer för öppna data både nationellt, på allt fler lärosäten och av forskningsfinansiärer, och flera av intervjupersonerna säger sig ha märkt av en utveckling mot ökad öppenhet runt data. Ännu verkar dock inte dessa policyer ha fått någon påtaglig betydelse för intervjupersonernas beslut runt datahantering – med undantag för när de formulerar forskningsansökningar. En intervjuperson påpekade också att kontrollen av att de policyer som redan finns också efterföljs ännu är bristande.¹⁴¹ Det går inte heller att komma ifrån att datatidskrifterna i sig kan vara ett försök av de vetenskapliga förlagen att skapa nya inkomstkällor för att stabilisera sin egen, allt mer ifrågasatta roll. Dataartikeln befäster också en kultur där vetenskapligt arbete mäts i antal publikationer, citeringar och genomslagsfaktorer, något som ingen av intervjupersonerna tog upp eller ifrågasatte. När det gäller delning av data hade det kanske i stället känts mer relevant att hitta metoder för att utvärdera till exempel vilka nya samarbeten och upptäckter delade data kan bidra till. En paradox är att forskningsfinansiärer i dag – med få undantag – premierar projekt som genererar nya data framför sådana som återanvänder redan existerande data (Borgman 2012). Kanske finns det på motsvarande sätt i

framtiden en risk för att projekt som generar data som är möjliga att dela och bevara kan komma att prioriteras framför projekt vars data inte (kan) delas. Det skulle i så fall kunna missgynna exempelvis små projekt i den långa svansen och forskningsprojekt med kvalitativa och/eller integritetskänsliga data.

7.2 Avslutande diskussion

Denna studie har visat hur ett urval forskare vid svenska universitet använder sig av dataartiklar eller i några fall delar sina data på andra sätt. Intervjuerna har även berört angränsande teman om bland annat samarbete, kulturer och praktiker runt datadelning, policyer för öppna data och motiv respektive hinder för delning av data. Frågorna är viktiga att studera ur samhällssynpunkt eftersom öppen tillgång till data kan bidra till nya upptäckter samtidigt som resurser sparas. Policyer för att påskynda utvecklingen mot öppen vetenskap införs därför av allt fler länder, lärosäten och forskningsfinansiärer. Även om forskningsbibliotekens arbete ligger utanför den här undersökningens avgränsningar kan resultaten vara av intresse för verksamheter som vill utveckla forskarstödet till att även omfatta stöd för exempelvis de mer praktiska aspekterna av datahantering och publicering av data (jag återkommer till detta under framtida forskningsfrågor, se avsnitt [7.2.1](#)).

Valet av teoretiskt ramverk och metod har haft både fördelar och nackdelar. Som metod passade kvalitativa metoder väl för att få svar på uppsatsens frågeställningar och gav forskarna möjlighet att beskriva såväl sitt arbete med datadelning som tankar om dataartiklens olika funktioner. Beslutet att intervjua forskare inom olika forskningsfält, med olika lång erfarenhet av forskning och olika typer av data baserades dels på en önskan att försöka skapa en så bred bild av användningen av dataartiklar som möjligt, dels på att urvalet av möjliga intervjupersoner var begränsat. Konsekvensen blev att det var svårt att uppnå mättnad i intervju svaren vad gäller de frågor i studien som rör specifika praktiker. Däremot anser jag att mättnad kan sägas ha uppnåtts när det gäller mer generella frågeställningar om delning av data och dataartiklens betydelse. Urvalet har också förhindrat en analys av de epistemiska kulturerna runt datadelning kopplat till ämnesområde. I stället har jag fått luta mig mot intervjupersonernas beskrivningar av sina arbetsplatser, kollegor, sina ämnesområden och data. Flera av intervjuerna bekräftar visserligen bilden av att det ändå inte finns någon enhetlig kultur runt datadelning på institutionsnivå utan att det snarare är runt smalare forskningsämnen eller till och med enskilda projekt som sådana praktiker uppstår. Med facit i hand hade det kanske varit mer intressant med ett urval av två grupper av intervjupersoner, som i Winnie Tams (2016; se avsnitt [2.1.3](#)) studie av forskare från två underdiscipliner inom geografi. Detta hade möjliggjort jämförelser och generaliserbarhet på ett helt annat sätt. Samtidigt hade det troligen inneburit svårigheter att hitta intervjupersoner som är verksamma vid svenska universitet, något som hade varit synd då det inte har gjorts så många intervjustudier om öppna data i en svensk kontext.

Som nämndes i metodavsnittet (se avsnitt [4.1](#)) hade det också varit önskvärt att komplettera intervjuerna med observationer på plats i forskarnas arbetsmiljö, något som tyvärr inte var möjligt inom ramen för denna uppsats med tanke på intervjupersonernas stora geografiska spridning. Även om själva arbetet med att skriva en dataartikel kanske inte hade varit så intressant att studera i sig hade forskarna kunnat visa de verktyg de använder för datainsamling, sina dataartiklar och hur de gick till väga med val av tidskrift, kontakt med förlaget, vilken artikelmall de använde, vilka data de delat, hur de förberedde dem för deponering etcetera. Utifrån enbart intervjuer har det varit svårare att få en

uppfattning om de data intervjupersonerna arbetar med och frågor och svar kom därför att handla om de sociala aspekterna av datadelning i större utsträckning än om de tekniska. Därmed var det också svårt att konsekvent tillämpa ett renodlat sociotekniskt perspektiv (där ju sociala och tekniska aspekter ses som svåra att separera från varandra) i analysdelen, som även den kom att få en tyngdpunkt på de sociala aspekterna. Observationer hade troligen kunnat bidra till en bättre balans i det avseendet. Jag har ändå haft glädje av det sociotekniska perspektivets fokus på samspelet mellan människor, teknik och praktiker liksom av Bijkers (2001) tankar om *interpretative flexibility* och hur stabiliseringen av en socioteknisk ensemble sker i förhandlingar mellan olika relevanta användargrupper.

7.2.1 Framtida forskningsfrågor

Arbetet med denna uppsats har väckt frågor om delning och publicering av data som inte har kunnat besvaras med enbart intervju som metod inom den begränsade tidsramen av en uppsats. Här följer därför några tankar om fortsatta forskningsfrågor och möjliga metoder för att studera dem:

Ett sätt att fördjupa kunskapen hade varit att följa ett eller flera forskningsprojekt från det att den första datahanteringsplanen görs, för att sedan observera arbetet med insamling, strukturering och bearbetning av data, och slutligen se hur data förbereds för delning och sedan deponeras. Eller att på motsvarande sätt följa processen när forskare återanvänder data som har delats av andra. Det borde också vara av intresse att jämföra ett forskningsfält där data delas ofta med ett annat där datadelning är ovanligt och dessutom intervjua både forskare som delar data och forskare som inte delar data inom båda fälten. Ett annat möjligt jämförelseområde är kvalitativa och kvantitativa data.

En enkätundersökning kan i större skala kartlägga svenska forskares attityder till delning av data och ge möjligheter till jämförelse med de internationella studier som gjorts. Dessutom kan en sådan undersökning ge information om huruvida det finns epistemiska kulturer kopplade specifikt till datadelning och om dessa kanske hänger ihop med en särskild typ av data snarare än till discipliner eller enskilda ämnesområden. I en enkätundersökning gör det också möjligt att undersöka om attityder till delning av data hänger samman med faktorer som ålder, kön och anställnings- och finansieringsform.

När det gäller just aktiviteterna runt specifikt dataartiklar tror jag dock, som jag redan har argumenterat för i metodavsnittet, att det kan vara svårt att fördjupa kunskaperna genom observationer eller enkätundersökningar. Då är det i stället mer relevant att göra bibliometriska studier av dataartiklar för att kartlägga förekomst fördelat på olika forskningsområden, länder och författare, göra ko-citeringsanalyser och även kartlägga ämnesord för att söka efter mönster runt exempelvis vilka typer av data som delas oftast. Eventuellt kan bibliometrisk mappning av dataartiklar också säga något om möjliga nya tvärvetenskapliga forskningsområden. Eftersom inte alla datatidskrifter finns indexerade i databaser som Scopus och Web of Science kan det dock innebära ett stort arbete att ett representativt urval av artiklar.

Datatidskrifter kan studeras med flera olika metoder. Ett sätt är att samla in material från datatidskrifters webbsajter där de beskriver syfte, ”scope”, open access-licenser, avtal runt upphovsrätt och riktlinjer för författare och granskare och göra en jämförande textanalys av detta material. Som nämndes i teoridelen (se avsnitt [3.2](#)) skulle även datatidskrif-

ten kunna ses som en socioteknisk ensemble och dess olika grupper av användare, tekniska förutsättningar och stabiliserande faktorer kan studeras utifrån detta perspektiv.

Slutligen är det av intresse att undersöka hur forskningsbiblioteken skulle kunna utveckla sitt arbete med forskarstöd när det gäller publicering av data. Publiceringsstrategier för *open access* är redan en fråga för biblioteken, och ibland kan de även kan erbjuda konkret stöd när det gäller publicering i Open Journal Systems. Detta arbete skulle kunna utvidgas till att även omfatta publicering av dataartiklar. Biblioteken skulle också – i samarbete med IT-personal – kunna erbjuda nya infrastrukturlösningar för att underlätta forskares arbete med andra sätt att publicering av data. Aktionsforskning i nära samarbete med professionen kan vara en givande metod för att utforska dessa frågor.

7.2.2 Delade data för fortsatta upptäckter

I uppsatsens inledning beskrevs hur nya upptäckter inom astronomi möjliggjorts genom att astronomer som var verksamma när forskaren själv ännu bara var barn råkat rikta sina teleskop åt exakt rätt håll och att deras data fortfarande var tillgängliga. Med tanke på de enorma mängder data som produceras idag och de allt mer avancerade metoder för dataanalys och samarbete som utvecklas är det hisnande att tänka på alla upptäckter som blivande forskare bland dagens förskolebarn kommer att kunna göra i framtiden. Detta förutsätter dock att de data som behövs verkligen delas, långtidsbevaras, och görs möjliga att hitta och återanvända. Långt ifrån alla data sparas idag och frågan är – just med tanke på de enorma datamängder som produceras – om det är rimligt eller ens möjligt att bevara alla data från alla forskare genom alla tider? Och om inte – hur kan vi då säkerställa att ”rätt” data bevaras och att det finns incitament för att detta görs? Detta är en viktig diskussion som troligen kommer att behöva föras länge än.

Källförteckning

Akers, K. G. (2016). Raising Researchers Awareness of Biomedical Data Journals to Promote Data Sharing. I Federer, L. (red) *The Medical Library Association Guide to Data Management for Librarians*. Lanham, Maryland: Rowman & Littlefield.

Alvesson, M. (2011). *Intervjuer – genomförande, tolkning och reflexivitet*. Malmö: Liber.

Anderberg, S. (2016). *Hantering av forskningsdata vid Stockholms universitet: Förslag till strategi och struktur*. Stockholm: Stockholms universitet. <http://su.diva-portal.org/smash/get/diva2:1073506/FULLTEXT01.pdf>

Bernhardt, B. R. (2016). Library Roles in Supporting Creation of Serials: Open Journal Systems at The University of North Carolina at Greensboro, *Serials Review*, 42(3) ss. 240-242. <https://doi.org/10.1080/00987913.2016.1204204>

Bijker, W. E. (1995). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. Cambridge, MA: MIT Press.

Bijker, W. E. (2001). Understanding Technological Culture through a Constructivist View of Science, Technology, and Society. I Mitcham, C., & Cutcliffe, S. H. (red): *Visions of STS: Counterpoints in Science, Technology, and Society Studies*. Albany: State University of New York Press.

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press.

Borgman, C. L. (2012). The conundrum of sharing research data, *Journal Of The American Society For Information Science & Technology*, 63(6), ss. 1059-1078. <https://doi.org/10.1002/asi.22634>

Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge, MA: MIT Press.

Callaghan, S. et al. (2012). Making Data a First Class Scientific Output: Data Citation and Publication by NERC's Environmental Data Centres, *International Journal Of Digital Curation*, 7(1), ss 107-113. <https://doi.org/10.2218/ijdc.v7i1.218>

Callaghan, S., Tedds, J., Lawrence, R., Murphy, P., Roberts, T. & Wilcox, W. (2014). Cross-Linking Between Journal Publications and Data Repositories: A Selection of Examples, *International Journal Of Digital Curation*, 9(1), ss 164-175. <https://doi.org/10.2218/ijdc.v9i1.310>

Candela, L., Castelli, D., Manghi, P., & Tani, A. (2015). Data journals: A survey. *Journal of the Association for Information Science and Technology*, 66(9), ss. 1747-1762. <https://doi.org/10.1002/asi.23358>

Chavan, V. & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* 12(15). <http://dx.doi.org/10.1186/1471-2105-12-S15-S2>

Chavan, V., Penev, L. & Hobern, D. (2016). Cultural Change in Data Publishing Is Essential. *BioScience*, 63(6) ss. 419-420. <https://doi.org/10.1525/bio.2013.63.6.3>

Chowdhury, G. (2014). OA policies and the sustainability of digital libraries of scholarly information. I Tuamsuk, K. et al. (red.) *The Emergence of Digital Libraries: Research and Practices*. (Lecture Notes in Computer Science, 8839.) Cham: Springer International Publishing, ss. 65-75. http://dx.doi.org/10.1007/978-3-319-12823-8_8

De Silva, P. U. K. & Vance, C. K. (2017). *Scientific Scholarly Communication – The Changing Landscape*. Cham: Springer International Publishing.

Elsevier & CWTS (2017). *Open Data: The Researchers Perspective*. https://www.elsevier.com/_data/assets/pdf_file/0004/281920/Open-data-report.pdf [18-03-02].

Eriksson-Zetterquist, U. & Ahrne, G. (2015). Intervjuer. I Ahrne, G. & Svensson, P. (red.) *Handbok i kvalitativa metoder*. Upplaga 2:2. Malmö: Liber, ss. 34-54.

Europakommissionen (u.å.). *Open Science (Open Access)*. <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access> [17-11-24].

Fagerström, E. (2016). Forskningsartiklar kostar miljoner – dränerar biblioteken. *Sydsvenskan*, 11 juni. <https://www.sydsvenskan.se/2016-06-11/forskningens-betalvaggar-tar-pa-bibliotekens-budgetar> [2018-04-26].

Forskningsdatautredningen (2017). *Personuppgiftsbehandling för forskningsändamål*. (SOU 2017:50). Stockholm: Utbildningsdepartementet.

Francke, H. (2008). *(Re)creations of scholarly journals : document and information architecture in open access journals*. Doktorsavhandling. Borås/Göteborg: Högskolan i Borås och Göteborgs universitet, Institutionen för Biblioteks- och informationsvetenskap/ Bibliotekshögskolan. <http://hb.divaportal.org/smash/get/diva2:876857/FULLTEXT01.pdf>

Fransson, J., du Toit, M., Kjellberg, S. & Tapia Lagunas, P. (2016). Tillgängliggörande av forskningsdata: nulägesbeskrivning samt rekommendationer för uppbyggnad av stöd till högskolans forskare. *Skrifter från Malmö högskolas bibliotek* 6(2016). <https://doi.org/10.24834/2043/20266>

Fransson, J., Kjellberg, S., Sedvall, K., Tengstam, O., Tapia Lagunas, P., du Toit, M., Widmark, J., Holmström, T. & Höie, H. (2017). På väg mot ett forskarstöd kring hantering och tillgängliggörande av forskningsdata: Slutrapport från biblioteksprojektet Tillgängliggörande av forskningsdata 2016. *Skrifter från Malmö högskolas bibliotek* 8(2017). <https://doi.org/10.24834/mahlib/2017/8>

Frohmann, B. (2004). *Deflating information: from science studies to documentation*. Toronto: University of Toronto Press.

GBIF (Global Biodiversity Information Facility) (u.å.). *Data papers – Getting scholarly recognition for your datasets*. <https://www.gbif.org/data-papers> [18.04.25].

Haider, J. (2017). Research data is a product of (its) time. I Kjellberg, S. & Moring, C. (red.), *Big data – small meaning and global discourses*. The Department of Arts and Cultural Sciences, Lund University. <https://medium.com/big-data-small-meaning-and-global-discourses/research-data-is-a-product-of-its-time-67e972aea9b7>

Haider, J. & Kjellberg, S. (2016). Data in the making: Temporal aspects in the construction of research data. I Rekers, J. V. & Sandell, K. (red.) *New Big Science In Focus: Perspectives on ESS And MAX IV*, (8), ss. 143-163. <http://lup.lub.lu.se/record/8852227>

Heidorn, P. B. (2008). Shedding Light on the Dark Data in the Long Tail of Science. *Library trends*, 57(2), ss. 280-299.

Hickson, S., Poulton, K. A., Connor, M., Richardson, J. & Wolski, M. (2016). Modifying researchers' data management practices. *IFLA Journal*, 42(4), ss. 253-265. <https://doi.org/10.1177/0340035216673856>

Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government, *Information Systems Management*. 29(4), ss. 258-268. <https://doi.org/10.1080/10580530.2012.716740>

Johnsson, M. & Lassi, M. (2016). *Hantering av forskningsdata vid fakulteterna inom Lunds universitet – en lägesbeskrivning hösten 2015*. Preprint, Universitetsbiblioteket, Lunds universitet, 2016. <http://lup.lub.lu.se/record/8725580>

Kennan, M. A. (2017). "In the eye of the beholder": knowledge and skills requirements for data professionals. *Information Research*, 22(4), ss. 1-21.

Kjellberg, S. (2010) *Forskarbloggar: Vetenskaplig kommunikation och kunskapsproduktion i bloggssfären*. Lund: Lunds universitet, Institutionen för kulturvetenskaper. <http://lup.lub.lu.se/record/1664543>

Kling, R., McKim, G. & King, A. (2003). A bit more to it: Scholarly communication forums as socio-technical interaction networks. *Journal of the American Society for Information Science and Technology*, 54(1), ss. 47-67. <https://doi.org/10.1002/asi.10154>

Klump, J. (2017). Data as Social Capital and the Gift Culture in Research. *Data Science Journal*, 16(14), ss. 1-8. <http://doi.org/10.5334/dsj-2017-014>

Knorr-Cetina, K. D. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Kratz, J. & Strasser, C. (2014). Data publication consensus and controversies [v3; ref status: indexed], *F1000research*, (3)94. <https://doi.org/10.12688/f1000research.3979.3>

Kratz, J. & Strasser, C. (2015). Researcher perspectives on publication and peer review of data, *Plos ONE*, 10(2), ss 1-21. <https://doi.org/10.1371/journal.pone.0117619>

Kruse, F. & Boserup Thestrup, J. (2018). I Kruse, F. & Boserup Thestrup, J. (red) *Research Data Management – A European Perspective*. Berlin/Borston: Walter de Gruyter GmbH, ss.1-8 .

Kungliga biblioteket (2018). *Sverige tar ställning för öppen tillgång – avtalet med Elsevier sägs upp* (Pressmeddelande den 16 maj).
http://www.mynewsdesk.com/se/kungliga_biblioteket/pressreleases/sverige-tar-staellning-foer-oeppen-tillgaang-avtalet-med-elsevier-saegs-upp-2508297 [2018-05-27].

Kvale, S. & Brinkmann, S. (2009). *Den kvalitativa forskningsintervjun*. Upplaga 2:1. Lund: Studentlitteratur.

Lalander, P. (2015). Intervjuer. I Ahrne, G. & Svensson, P. (red.) *Handbok i kvalitativa metoder*. Upplaga 2:2. Malmö: Liber, ss. 93-112.

Lawrence, B., Jones, C., Matthews, B., Pepler S. & Callaghan, S. (2011), Citation and Peer Review of Data: Moving Towards Formal Data Publication, *International Journal Of Digital Curation*, 6(2), ss 4-37. <https://doi.org/10.2218/ijdc.v6i2.205>

Lassi, M., Johnsson, M., & Golub, K. (2016). Research data services, *IFLA Journal*, 42(4), ss. 266-277. <https://doi.org/10.1177/0340035216671963>

Leonelli, S. (2016). *Data-centric biology: a philosophical study*. Chicago; London: The University of Chicago Press.

Levy, D. M. (2003). Documents and Libraries: A Sociotechnical Perspective. I Peterson Bishop, A., Van House, N. A. & Battenfield, B. P. (red): *Digital Library Use: Social Practice in Design and Evaluation*. Cambridge, MA: MIT Press, ss. 25-42.

Mauthner, N., S. & Parry, O. (2013). Open Access Digital Data Sharing: Principles, Policies and Practices, *Social Epistemology*, 27(1), ss. 47-67.
<http://dx.doi.org/10.1080/02691728.2012.760663>

Meyer, E. (2018). Big Data is Transforming How Astronomers Make Discoveries: the next game-changer is likely lurking in the data we already have – but it will take scientists years to uncover it. *Smithsonian.com*, 15 maj.
<https://www.smithsonianmag.com/science-nature/next-big-discovery-astronomy-scientists-probably-found-it-years-ago-they-dont-know-it-yet-180969073/> [2018-05-16].

Nelhans, G. & Eklund, P. (2013). *Resursfördelningsmodeller på bibliometrisk grund vid ett urval svenska lärosäten*. (Rapport 2015:30) Borås: Vetenskap för profession.
<http://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-21>

Park, J.-H. (2008). The Relationship between Scholarly Communication and Science and Technology Studies (STS), *Journal of Scholarly Publishing*, 39(3), ss. 257-274.
<http://dx.doi.org/10.3138/jsp.39.3.257>

- Park, H. & Wolfram, D. (2017). An examination of research data sharing and re-use: implications for data citation practice, *Scientometrics*, 111(1) ss. 443-461.
<http://dx.doi.org/10.1007/s11192-017-2240-2>.
- Parsons, M. A., Duerr, R. & Minster, J.-B. (2010). Data Citation and Peer Review, *Eos, Transactions, American Geophysical Union Volume*, 91(34), ss. 297–298.
<https://doi.org/10.1029/2010EO340001>
- Piwowar, H. A. & Vision, T.J. (2013) Data reuse and the open data citation advantage. *PeerJ*, (1)175 <https://doi.org/10.7717/peerj.175>
- Pryor, G. (Red.). (2012). Why manage research data? I Pryor, G. (red.) *Managing Research Data*. London: Facet Publishing, ss. 1-15.
- Rennstam, J. & Wästerfors, D. (2015). Att analysera kvalitativt material. I Ahrne, G. & Svensson, P. (red.) *Handbok i kvalitativa metoder*. Upplaga 2:2. Malmö: Liber, ss. 221-236.
- Robertson, W. & Simser, C. (2013). Managing E-Publishing: Perfect Harmony for Serialists, *The Serials Librarian*, 64(1-4), ss. 118-128.
<https://doi.org/10.1080/0361526X.2013.760399>
- Ryan, G. W. & Bernard, H. R. (2003). Techniques to Identify Themes. *Field Methods*, 15(1), ss. 85–109. <https://doi.org/10.1177/1525822X02239569>
- Suber, P. (2012). *Open access*. Cambridge, MA: MIT Press.
- Svensson, P. (2015). Teorins roll i kvalitativ forskning. I Ahrne, G. & Svensson, P. (red.) *Handbok i kvalitativa metoder*. Upplaga 2:2. Malmö: Liber, ss. 208-219.
- Svensson, P. & Ahrne, G. (2015). Att designa ett kvalitativt forskningsprojekt. I Ahrne, G. & Svensson, P. (red.) *Handbok i kvalitativa metoder*. Upplaga 2:2. Malmö: Liber, ss. 17-31.
- Tam, W. (2016). *Discipline and research data in geography*. Loughborough: Loughborough University, The Centre for Information Management.
<https://dspace.lboro.ac.uk/2134/26997>
- Tam, W., Fry, J. & Proberts, S. (2014). The Disciplinary Shaping of Research Data Management Practices. In *iConference 2014 Proceedings*, ss. 721–728.
<https://doi.org/10.9776/14338>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A., Wu, L., Read, E., Manoff, M. & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *Plos ONE* 6(6), ss 1-21. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, C., Sandusky, R., Allard, S. & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library And Information Science Research*, 36, ss. 84-90. <http://dx.doi.org/10.1016/j.lisr.2013.11.003>

Tracz, V. & Lawrence, R. 2016, Towards an open science publishing platform, *F1000research*, 5, ss. 1-10. <https://doi.org/10.12688/f1000research.7968.1>

Trost, J. (2014). *Att skriva uppsats med akribi*, upplaga 4:1. Lund: Studentlitteratur.

Utbildningsdepartementet (2016). *Kunskap i samverkan: för samhällets utmaningar och stärkt konkurrenskraft* (Regeringens proposition 2016/17:50). Stockholm: Regeringskansliet.

Van den Eynden, V. & Bishop, L. (2014) *Incentives and motivations for sharing research data, a researcher's perspective*. UK Data Archive (rapport). University of Essex: Knowledge Exchange. http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf [2018-03-16].

Van House, N. A (2003a). Digital Libraries and Collaborative Knowledge construction. I Van House, N. A., Peterson Bishop, A. & Pfeil Battenfield, B. (red.): *Digital library use: Social practice in design and evaluation*. Cambridge, MA: MIT Press, ss. 271-295.

Van House, N. A. (2003b). Science and Technology Studies and Information Studies. *Annual Review of Information Science and Technology* vol. 38(1), ss. 3-86. <https://doi.org/10.1002/aris.1440380102>

Vetenskapsrådet (2015). *Öppen tillgång (Open access) till forskningsdata: en omvärldsbeskrivning 2014*. Stockholm: Vetenskapsrådet. <https://publikationer.vr.se/produkt/oppen-tillgang-open-access-till-forskningsdata-en-omvarldsbeskrivning-2014/> [18-01-12]

Wallis, J., Rolando, E. & Borgman, C. L. (2013), If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology, *Plos ONE*, 8(7) ss. 1-17. <https://doi.org/10.1371/journal.pone.0067332>

Whyte, A. & Pryor, G. (2011): Open science in practice: researcher perspectives and participation. *International journal of digital curation* 1(6), ss. 199-213. <https://doi.org/10.2218/ijdc.v6i1.182>

Ya-Ning, C. (2016). An analysis of characteristics and structures embedded in data papers: a preliminary study, *Libellarium: Journal For The Research Of Writing, Books & Cultural Heritage Institutions*, 9(2) ss. 145-156.

Zarate, O., Brody, J., Brown, P., Ramirez-Andreotta, M., Perovich, L. & Matz, J. (2016). *Balancing Benefits and Risks of Immortal Data*. (Rapport 46/1) Hastings Center Report, ss. 36-45. <https://doi.org/10.1002/hast.523>

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., and Alibaks, R. S. (2012). Socio-technical Impediments of Open Data. *Electronic Journal of e-Government*, 10(2), ss 156-172.

Bilaga 1 · Intervjuförfrågningar

Hej!

Mitt namn är Karin Sandqvist och jag studerar biblioteks- och informationsvetenskap på ABM-masterprogrammet vid Lunds universitet. Under den här terminen arbetar jag med min masteruppsats som handlar om *data papers*, korta artiklar som beskriver dataset.

I uppsatsen vill jag undersöka dels vilken roll sådana dataartiklar kan spela för att möjliggöra återanvändning av öppna forskningsdata inom olika forskningsområden, dels vilken betydelse det har att dataartiklarna kan granskas, citeras och vara en del av meriteringssystemet på samma sätt som en konventionell vetenskaplig artikel.

Syftet är att öka kunskapen om dataartiklars funktion i förhållande till öppen vetenskap och öppna data. Resultaten kan vara av intresse för forskare som vill dela sina egna data och kunna återanvända andras, men även mer generellt inom forskning kring vetenskaplig kommunikation och öppen vetenskap.

Under mina sökningar i Web of Science, Scopus och i Svensk nationell datatjänsts arkiv har jag funnit att du är en av de forskare i Sverige som har skrivit eller citerat minst en dataartikel eller på andra sätt aktivt delat dina data och därför skulle jag gärna vilja intervjua dig om dina erfarenheter.

Jag planerar att genomföra intervjuerna under vecka 8-10 och jag uppskattar att de kommer att ta cirka 30 minuter. Med ditt godkännande görs en ljudinspelning av intervjun.

Vill du delta i undersökningen?

Tveka inte att höra av dig till mig eller min handledare om du har några frågor!

Vänliga hälsningar

Karin Sandqvist
mejladress
telefonnummer

Handledare:
Jutta Haider
mejladress

Bilaga 2 · Intervjuguide

Typ av data

1. Vilken sorts data har du delat och vilket/vilka program har du använt?
2. Var finns datasetet/dataseten?
3. Delar du alla dina data eller finns det data du har valt att inte dela och i så fall varför?
4. Har du haft några betänkligheter runt den data du har delat (till exempel av etiska skäl)?
5. Hur gick det till rent tekniskt när du skulle dela dina data, stötte du på några svårigheter?

Dataartikel

6. Hur fick du idén att skriva en dataartikel?
7. Hur gick det till rent praktiskt (fanns det t ex en mall att utgå från)?
8. Vad är skillnaden jämfört med att skriva en vanlig vetenskaplig artikel?
9. Vad är skillnaden jämfört med att beskriva ett dataset i metadata eller under metoddelen i en "vanlig" artikel?
10. Har du delat data på andra sätt, hur skiljer det sig i så fall från att skriva en dataartikel?

Citering av dataartikel (alternativa frågor till forskare som bara citerat har dataartikel)

6. Hur fick du kännedom om dataartikel/dataartiklar som du har citerat?
7. Tittade du även på själva datasetet eller räckte informationen som fanns i dataartikeln?
8. Är det någon skillnad att citera en dataartikel jämfört med att citera en traditionell vetenskaplig artikel?
9. Är det någon skillnad att citera en dataartikel jämfört med att citera data som delats på andra sätt?
10. Vad är den största skillnaden mellan en dataartikel och beskrivningen av ett dataset som metadata eller under metoddelen i en "vanlig" artikel?

Granskning och citering

11. Blev din dataartikel peer reviewad och vilken betydelse har det i så fall?
12. Har din dataartikel citerats och vilken betydelse har det i så fall?
13. Vilken betydelse har det att din dataartikel kan bli en del av meriteringssystemet på samma sätt som en vanlig vetenskaplig artikel?
14. Har du själv citerat din dataartikel (eller andras dataset)?
15. Hur hittade du i så fall din dataartikeln/datasetet?

Mekanismer runt delning av forskningsdata

16. Vad tror du är de största hindren för forskare att dela sina data?
17. Finns det något som skulle ha underlättat för dig att dela dina data?
18. Har någon policy (nationell, finansiärs, det egna lärosätets, förlags) haft betydelse för att du delat forskningsdata?
19. Vilken kultur runt delning av forskningsdata inom ditt forskningsområde?
20. Skulle du kunna tänka dig att skriva fler dataartiklar?

Bilaga 3 · Informerat samtycke

Dataartiklars betydelse för delning av öppna forskningsdata

Bakgrund och syfte

Denna intervjustudie görs som en del av arbetet med en masteruppsats i biblioteks- och informationsvetenskap. Uppsatsen handlar om *data papers*, korta artiklar som beskriver dataset, och vilken roll sådana dataartiklar kan spela för att möjliggöra återanvändning av öppna forskningsdata inom olika forskningsområden och i jämförelse med andra sätt att dela data. Syftet är att öka kunskapen om dataartiklars funktion i förhållande till öppen vetenskap.

Deltagande och genomförande av intervjustudien

Det empiriska materialet till uppsatsen kommer huvudsakligen att samlas in genom intervjuer med forskare i Sverige som har skrivit, citerat och/eller granskat dataartiklar eller på andra sätt aktivt delat sina data. Med varje informant görs en intervju som uppskattas ta cirka 30 minuter och där ljudet – med informantens godkännande – spelas in.

Risker och fördelar

Deltagande i undersökningen innebär inga kända risker. Resultaten kan vara intressanta för forskare som vill tillgängliggöra sina egna data och kunna återanvända andras, men även inom forskning kring öppen vetenskap, öppna data och vetenskaplig kommunikation.

Konfidentialitet och hantering av data

Informantens namn kommer inte att anges i uppsatsen och inte heller vid vilket lärosäte informanten arbetar. Däremot kommer det att framgå vilken befattning informanten har (till exempel professor, docent, doktorand), och det kan även vara relevant att berätta inom vilket forskningsområde informanten är verksam. Detta kan, tillsammans med det faktum att så få forskare i Sverige har publicerat dataartiklar, delvis begränsa möjligheten till konfidentialitet. Intervjumaterial som samlas in i studien avidentifieras och informanternas namn kommer att ersättas med koder. Kodnyckeln bevaras så att obehöriga inte kan komma åt den.

Frivilligt deltagande

Deltagande i undersökningen är frivilligt och informanter kan när som helst och utan förklaring välja att avbryta sin medverkan.

Ansvar

Uppsatsens handledare är Jutta Haider, docent och universitetslektor i biblioteks- och informationsvetenskap vid institutionen för kulturvetenskaper, Lunds universitet.

[Mejladress](#)

Kontakt

Karin Sandqvist, student vid masterprogrammet i ABM vid Lunds universitet.

Telefonnummer · [Mejladress](#)

Formulär för informerat samtycke

Deltagande i undersökning:

Jag har tagit del av informationen rörande undersökningen. Jag har fått tillfälle att ställa frågor om undersökningen och de har blivit besvarade. Jag samtycker till att delta i denna undersökning.

Deltagarens underskrift: _____

Namnförtydligande: _____

Datum (dag/månad/år): _____

Intervjuarens underskrift: _____

Namnförtydligande: _____

Datum: _____

Appendix · Forskarna och intervjuerna

1. Professor, sociologi. Har delat data i SND:s arkiv. Intervjuades via Skype den 22 februari, 34 minuter.
2. Externfinansierad forskare (har doktorerat), språkvetenskap. Medförfattare till en dataartikel samt har delat data i SND:s databas. Intervjuades via Skype den 26 februari, 32 minuter.
3. Universitetslektor, naturgeografi. Har skrivit två dataartiklar och håller på med en tredje. Intervjuades via Skype den 27 februari, 29 minuter.
4. Doktorand, ekosystemvetenskap. Har skrivit en dataartikel och citerat en anna dataartikel. Intervjuades 28/2, 35 minuter.
5. Professor, statsvetenskap. Har skrivit, citerat och granskat flera dataartiklar samt delat data i SND:s databas. Intervjuades den 5 mars, 58 minuter.
6. Doktorand, klinisk medicin. Har skrivit en dataartikel. Intervjuades via Skype den 16 mars, 30 minuter.
7. Doktorand, ekologi, miljö och botanik. Har citerat en dataartikel. Svarade på frågor via mejl den 28 februari och på en uppföljande fråga den 2 mars.
8. Forskare med postdok-tjänst, ekonomisk demografi. Har skrivit en dataartikel. Svarade på frågor via mejl den 15 mars och på en uppföljande fråga den 21 mars.

Samtliga forskare arbetar vid större svenska statliga universitet.

Appendix

· Nyckel till citaten

¹ Intervju 1.
² Intervju 2.
³ Intervju 1.
⁴ Intervju 7.
⁵ Egen nyckel.
⁶ Egen nyckel.
⁷ Egen nyckel.
⁸ Intervju 3.
⁹ Egen nyckel
¹⁰ Egen nyckel.
¹¹ Intervju 8.
¹² Egen nyckel.
¹³ Egen nyckel.
¹⁴ Egen nyckel.
¹⁵ Egen nyckel.
¹⁶ Egen nyckel.
¹⁷ Intervju 8.
¹⁸ Intervju 7.
¹⁹ Intervju 2.
²⁰ Intervju 6.
²¹ Intervju 2, 6.
²² Intervju 6.
²³ Intervju 2.
²⁴ Intervju 8.
²⁵ Intervju 3.
²⁶ Intervju 1.
²⁷ Intervju 5.
²⁸ Intervju 5.
²⁹ Intervju 5.
³⁰ Intervju 6.
³¹ Intervju 6.
³² Intervju 5.
³³ Intervju 3.
³⁴ Intervju 3.
³⁵ Intervju 8.
³⁶ Intervju 1.
³⁷ Intervju 8.
³⁸ Intervju 6.
³⁹ Intervju 2.
⁴⁰ Intervju 6, 7.
⁴¹ Intervju 7.
⁴² Intervju 6.
⁴³ Intervju 7.
⁴⁴ Intervju 7.
⁴⁵ Intervju 1.

⁴⁶ Intervju 4.
⁴⁷ Intervju 4.
⁴⁸ Intervju 3.
⁴⁹ Intervju 2.
⁵⁰ Intervju 4.
⁵¹ Intervju 3.
⁵² Intervju 3.
⁵³ Intervju 2.
⁵⁴ Intervju 5.
⁵⁵ Intervju 3.
⁵⁶ Intervju 5.
⁵⁷ Intervju 5.
⁵⁸ Intervju 5.
⁵⁹ Intervju 5.
⁶⁰ Intervju 8.
⁶¹ Intervju 5.
⁶² Intervju 3.
⁶³ Intervju 2.
⁶⁴ Intervju 4.
⁶⁵ Intervju 2, 3, 4.
⁶⁶ Intervju 4.
⁶⁷ Intervju 7.
⁶⁸ Intervju 7.
⁶⁹ Intervju 8.
⁷⁰ Intervju 2.
⁷¹ Intervju 3.
⁷² Intervju 1.
⁷³ Intervju 3.
⁷⁴ Intervju 4.
⁷⁵ Intervju 3.
⁷⁶ Intervju 6.
⁷⁷ Intervju 8.
⁷⁸ Intervju 4.
⁷⁹ Intervju 6.
⁸⁰ Intervju 3.
⁸¹ Intervju 3.
⁸² Intervju 3.
⁸³ Intervju 4.
⁸⁴ Intervju 6.
⁸⁵ Intervju 7.
⁸⁶ Intervju 5.
⁸⁷ Intervju 4.
⁸⁸ Intervju 3.
⁸⁹ Intervju 3.
⁹⁰ Intervju 8.
⁹¹ Intervju 6.
⁹² Intervju 4.
⁹³ Intervju 4.

⁹⁴ Intervju 5.
⁹⁵ Intervju 3.
⁹⁶ Intervju 3.
⁹⁷ Intervju 5.
⁹⁸ Intervju 5.
⁹⁹ Intervju 5.
¹⁰⁰ Intervju 3.
¹⁰¹ Intervju 8.
¹⁰² Intervju 4.
¹⁰³ Intervju 7.
¹⁰⁴ Intervju 5.
¹⁰⁵ Intervju 5.
¹⁰⁶ Intervju 4.
¹⁰⁷ Intervju 5.
¹⁰⁸ Intervju 4.
¹⁰⁹ Intervju 5.
¹¹⁰ Intervju 4.
¹¹¹ Intervju 3.
¹¹² Intervju 3.
¹¹³ Intervju 3.
¹¹⁴ Intervju 5.
¹¹⁵ Intervju 3.
¹¹⁶ Intervju 6.
¹¹⁷ Intervju 5.
¹¹⁸ Intervju 8.
¹¹⁹ Intervju 8.
¹²⁰ Intervju 6.
¹²¹ Intervju 3.
¹²² Intervju 1, 2.
¹²³ Intervju 1.
¹²⁴ Intervju 1.
¹²⁵ Intervju 4, 6, 8.
¹²⁶ Intervju 4.
¹²⁷ Intervju 5.
¹²⁸ Intervju 3.
¹²⁹ Intervju 8.
¹³⁰ Intervju 5.
¹³¹ Intervju 7.
¹³² Intervju 8.
¹³³ Intervju 1.
¹³⁴ Intervju 2.
¹³⁵ Intervju 2 & 5.
¹³⁶ Intervju 2, 3, 4, 6, 8.
¹³⁷ Egen nyckel.
¹³⁸ Intervju 3.
¹³⁹ Intervju 8.
¹⁴⁰ Intervju 3 & 4.
¹⁴¹ Intervju 3.