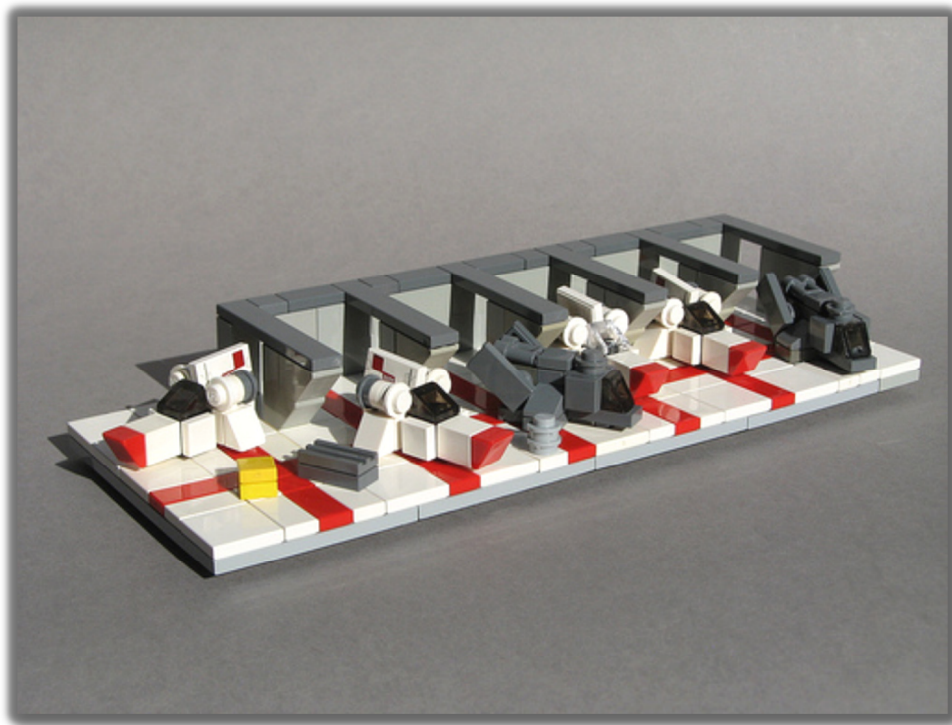




Projektet Forskningsdata inom humaniora och konstnärliga vetenskaper – open access?

Urban Andersson
Iris Alfredsson
Sofia Arvidsson
Jörgen Eriksson
Birger Jerlehag
David Lawrence
Jessica Lindholm

Projektrapport till Kungl. biblioteket,
Programmet för OpenAccess.se
April 2011

Tack

Vi tackar våra forskare som gav sig tid att träffa oss och besvara våra intervjufrågor, och som genom engagerande diskussion gav oss inblickar i problematiken med forskningsdata.

Vi tackar också Kungl. biblioteket för finansiellt stöd till projektet.

Göteborg, Linköping, Lund och Malmö den 20 april 2011

Innehållsförteckning

1. Sammanfattning, slutsatser & vision	3
2. Inledning	5
2.1 Projektets syfte	5
2.2 Projektinformation	6
2.2.1 Deltagare	6
2.2.2 Projektledning	6
2.2.3 Uppdrag och finansiering	6
2.2.4 Projektrapportering	6
2.3 Förstudie	6
2.4 Internationell utblick	7
2.5 Aktuella verksamheter	8
2.5.1 Göteborgs universitet	8
2.5.2 Linköpings universitet	9
2.5.3 Lunds universitet	10
2.5.4 Malmö högskola	10
2.5.5 Svensk nationell datatjänst (SND)	10
3. Genomförande	11
4. Intervjuer, sammanställning	12
4.1 Forskningsprojekt	12
4.2 Fråga 1 – Insamling och bearbetning av forskningsdata	12
4.3 Fråga 2 – Lagringsmedium	13
4.4 Fråga 3 – Publicering av resultaten från forskningsprojekten	14
4.5 Fråga 4 – Tillgängliggörande av forskningsdata	15
4.6 Fråga 5 – Återanvändning av forskningsdata	17
4.7 Fråga 6 – Spridning genom länkning mellan publikation och forskningsdata	18
4.8 Fråga 7 – Hjälp	18
4.9 Fråga 8 – Deltagande i projektet	18
5. Typer av data, problem, utmaningar och lärdomar.....	20
6. Systemval	23

APPENDIX

1. Sammanfattning, slutsatser & vision

Vad är forskningsdata?

I OECD Principles and Guidelines for Access to Research Data from Public Funding definieras forskningsdata som:

“factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.”¹

Traditionellt härrör begreppet forskningsdata främst från samhällsvetenskaperna och syftar primärt på text och numeriska data.

Forskningsdata inom humaniora och konstnärliga vetenskaper är per definition ofta av helt annan karaktär, både vad avser det fysiska formatet och hur det hanteras och används inom forskningen.

Projektet Forskningsdata inom humaniora och konstnärliga vetenskaper – open access? bygger på en förstudie som genomfördes av universitetsbiblioteken i Göteborg och Lund under 2008. I den föreslogs bland annat att ett projekt borde begränsa sig till en eller ett par ämnesdiscipliner, då förutsättningar och behov kan skilja sig markant. Den föreslog vidare att man som metod skulle välja djupintervju, i stället för enkätundersökning, vilket var den metod som användes i förstudien.

Ett 20-tal intervjuer med forskare har utförts inom ramen för projektet.

Några kärnfrågor återkom i de allra flesta intervjuer.

Bland annat när i forskningsprocessen som datafrågorna aktualiseras, vem som äger rätten till forskningsdata, liksom vad det faktiska värdet av att tillhandahålla data egentligen är.

Ett problem som nämns av många är att svårigheterna med att i efterhand försöka dokumentera och hantera forskningsdata. I synnerhet då data ofta är av såväl varierande format som varierande kvalitet och mer eller mindre svårt att hantera på ett meningsfullt sätt. Man påpekar att detta i stället måste ske kontinuerligt, parallellt med det övriga forskningsarbetet.

Samma erfarenheter framkommer när det gäller lagringen av data.

Idag lagras data ofta på en mängd olika sätt, i många fall även inom ett och samma forskningsprojekt. Som exempel nämns egna hemsidor, olika former av portabla lagringsmedia, institutionens servrar, externa portaler och nätverk, sociala nätverk som YouTube. Man nämner även alternativen “i huvudet” eller – ännu värre – “inte alls”. Datas beständighet är en stark osäkerhetsfaktor, oavsett hur dessa sparas.

¹ <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

Här efterlyser man befintliga, enhetliga system, liksom stöd – både tekniskt och policymässigt, vad beträffar upphovsrätt och tillgänglighet. I viss mån finns detta idag, bland annat genom Svensk nationell datatjänsts försorg, men medvetenheten om detta är låg.

Generellt så ser vi dock en i huvudsak positiv attityd till att lagra och tillgängliggöra forskningsdata open access.

Vad som idag avhåller forskare från att tillgängliggöra forskningsdata, utöver vad som nämnts ovan, kan i grova drag delas in i två kategorier – tekniska svårigheter och motivation.

Forskningsdata inom humanistiska och konstnärliga vetenskaper utgörs av en mycket stor mängd dataformat. Vissa är fria och/eller stöds av befintliga och allmänt tillgängliga verktyg. Andra är proprietära och kräver specialprogramvara, vissa i så hög grad att de till och med förutsätter unik hårdvara för att kunna utnyttjas på ett meningsfullt sätt. Här är naturligtvis även formatens livslängd av stor betydelse. Användandet av, i möjligaste mån, öppna och arkivbeständiga format, liksom möjligheten att kunna erbjuda konvertering av data är viktigt, parallellt med användandet av erkända standarder.

En annan kritisk fråga i sammanhanget är själva datamängden. Lagring av film- och ljudinspelningar i digital form kräver mycket stora serverutrymmen. Detta kan givetvis också försvåra tillgängliggörandet. Detta ställer höga krav på såväl system som systemkompetens hos berörda parter.

Av intervjumaterialet framgår att det finns viss osäkerhet vad beträffar upphovsrätten

Vi har också stött på en rädsla för att felaktigheter i det ingående datat skall upptäckas av forskarvärlden, liksom en rädsla för stöld. Vad det sistnämnda beträffar så skall tilläggas att vi i studierna av internationella projekt har mötts av åsikten att denna rädsla ofta är obefogad, liksom att det även kan vara positivt för den egna forskningen om andra forskare får tillgång till ingående forskningsdata.

Integritetsproblem och frågor som rör anonymisering av forskningsdata, vilket är vanligt förekommande inom den samhällsvetenskapliga forskningen, förekommer i viss mån även inom humanistisk och konstnärlig forskning, om än inte i lika hög grad.

Slutligen stöter vi i några av intervjuerna på en något bristande motivation, att man helt enkelt inte ser fördelarna med att tillhandahålla sina forskningsdata. Även här ser vi att såväl ändamålsenliga system som kompetent stöd är oerhört viktigt.

I synnerhet när det gäller den konstnärliga forskningen så är även presentationsformaten av stor vikt, i vissa fall till och med helt avgörande. En åsikt som framkommer är att dagens publiceringssystem, vilka ju främst är konstruerade för att hantera textmaterial, inte är helt anpassade till dessa discipliners krav på utformning och design.

Det samma gäller även befintliga metadatamodeller, som bör ses över och utökas för att tillgodose de krav som konstnärlig forskningsdata, liksom forskningsdata i allmänhet ställer.

För att motivera forskare att lagra och tillgängliggöra sina forskningsdata open access måste vinsterna tydliggöras.

Vision

Mot bakgrund av ovanstående har vi inom projektet utformat en vision av hur vi skulle vilja se att forskningsdata inom berörda ämnesområden hanteras i framtiden.

Denna vision omfattar:

- att en detaljerad, realistisk och hållbar plan för lagring och bevaring, liksom för det framtida tillgängliggörandet av forskningsdata utarbetas och dokumenteras redan vid ett forskningsprojektets start.
- att forskningsbiblioteken i samarbete med andra aktörer, såsom universitets- och högskolearkiven, Kungliga biblioteket och Svensk nationell datatjänst tillhandahåller såväl kompetensmässigt stöd som fungerande teknisk infrastruktur för hantering av forskningsdata. Denna hantering omfattar beskrivning, bearbetning, lagring och tillgängliggörande.
- uppbyggandet av en infrastruktur, omfattande system, rutiner och standarder för insamling, beskrivning och lagring av forskningsdata samt sammanlänkning med publikation finns. I denna infrastruktur ingår även system för att på ett persistent sätt kunna identifiera data sets, samt möjliggöra delning av information och data mellan olika system.

2. Inledning

2.1 Projektets syfte

Projektet syftar till att öka tillgängligheten av forskningsdata inom humaniora och konstnärliga vetenskaper och skall försöka belysa följande frågeställningar:

- Var och hur skall forskningsdata lagras?
- Vilka delar kan publiceras som open access?
- Hur bör koppling ske mellan öppna arkiv och Svensk Nationell Datatjänst?
- Hur bör koppling ske mellan publikation och tillhörande forskningsdata?

2.2 Projektinformation

2.2.1 Deltagare

Deltagande lärosäten är universiteten i Göteborg, Linköping och Lund, Malmö högskola samt Svensk Nationell Datatjänst (SND) vid Göteborgs universitet.

2.2.2 Projektledning

Projektet leds av Göteborgs universitetsbibliotek. Projektledare är Urban Andersson, som vid projektets start ersatte den ursprungligen tänkta projektledaren, Carin Björklund.

2.2.3 Uppdrag och finansiering

Projektet är finansierat med 500,000 kronor av Kungl. Biblioteket, inom ramen för programmet OpenAccess.se, som drivs i samarbete med KK-stiftelsen, Kungl. Vetenskapsakademien, Riksbankens Jubileumsfond, SUHF och Vetenskapsrådet.

2.2.4 Projektrapportering

Slutrapportering sker i och med avlämnandet av denna rapport. En lägesrapport lämnades i juni 2010.

Projektet har även redovisats vid Open Access-dagarna i Uppsala 2009, respektive Lund 2010.

2.3 Förstudie

Under 2007 genomfördes en förstudie, i form av ett projekt finansierat av OpenAccess.se, vid universiteten i Göteborg, Lund och vid Sveriges Lantbruksuniversitet.

Resultatet av denna publicerades i en slutrapport med titeln "Forskningsdata i öppna arkiv och universitetsarkiv: en förstudie vid Göteborgs universitet, Lunds universitet och Sveriges Lantbruksuniversitet"²

Denna studie, som främst genomfördes i form av en ämnesövergripande enkätundersökning, visade på ett intresse för frågan bland forskarna, men också på viss osäkerhet. Man konstaterade också att svaren inte skiljde sig nämnvärt från de resultat som presenterats i internationella studier.

Man såg även att forskare inom vissa områden inom de humanistiska och konstnärliga vetenskaperna är på väg att knappa in på det försprång naturvetare hittills haft.

² <http://hdl.handle.net/2077/7379>

2.4 Internationell utblick

I projektet har vi även studerat ett antal olika internationella initiativ, av vilka kan nämnas:

CESSDA

Council of European Social Science Data Archives (CESSDA)³, är en paraplyorganisation för ett 20-tal europeiska dataarkiv där Svensk Nationell Datatjänst (SND) är en av medlemmarna.

Just nu planerar CESSDA att bygga upp en fastare infrastruktur, CESSDA ERIC. Starten är beräknad till 2012. Sverige kommer att ingå som medlem, med SND som svensk ”service provider”.

CESSDA tillhandahåller också en egen söktjänst, där metdata över forskningsdata som tillhandahålls av medlemsorganisationerna registreras och görs publikt sökbar.

DDI

Data Documentation Initiative Alliance (DDI)⁴ är ett internationellt initiativ, med syfte att skapa en XML-standard för beskrivning och utväxling av forskningsdata, främst inom samhällsvetenskaperna, DDI Lifecycle⁵.

DDI är en medlemsstyrd organisation och Svensk Nationell Datatjänst är en av dess medlemmar.

DANS

Data Archiving and networked services (DANS)⁶ är holländsk institution som, sedan 2005, ansvarar för arkivering och tillgängliggörande av forskningsdata inom samhällsvetenskap och humaniora i Nederländerna.

DANS har ansvaret för framtagandet av Data Seal of Approval⁷ - totalt 16 st riktlinjer för kvalitetssäkring av forskningsdata inom dessa områden. I dagsläget har fem elektroniska arkiv erhållit Data Seal of Approval.⁸

DANS tillhandahåller även ett system – Electronic Archiving System, EASY⁹ – för deponering av forskningsdata, liksom framsökning av data.

Forskningsdata i EASY kan, i förekommande fall (Open access) laddas ner online av slutanvändare efter inloggning. Systemet använder URN för persistent identifiering av ingående objekt.

Här ingår också en växande mängd retrospektiv digitalisering och registrering/lagring av forskningsdata, genomfört enligt metoden ADA¹⁰ (Academic Digital Archiving).

³ <http://www.cessda.org>

⁴ <http://www.ddialliance.org/>

⁵ <http://www.ddialliance.org/Specification/DDI-Lifecycle/3.1>

⁶ <http://www.dans.knaw.nl/en>

⁷ <http://www.datasealofapproval.org/>

⁸ <http://assessment.datasealofapproval.org/seals/>

⁹ DANS EASY, <https://easy.dans.knaw.nl/dms>

¹⁰ <http://www.dans.knaw.nl/en/content/categorieen/diensten/retro-archiving-data-ada-method>

En metod som har utarbetats inom institutet och som omfattar samtliga delar av denna process, såsom inventering, urval, tillägg av metadata och lagring.

DANS är verksam inom Royal Netherlands Academy of Arts and Sciences (KNAW).

EDINA

Edinburgh Data Share (EDINA)¹¹ är ett brittiskt, nationellt akademisk data center, ursprungligen initierat av JISC, 1996, med uppdrag att, bland annat, tillhandahålla tjänster för hantering av forskningsdata för hela det brittiska akademiska samfundet, omfattande samtliga forskningsområden och med fokus på delning av data. Bland de tjänster som tillhandahålls av EDINA kan nämnas ShareGeo Open¹², en tjänst för att registrera, söka och dela geospatiella datasets.

Under 2007-2009 ledde EDINA projektet JISC-UK DataShare¹³, ett projekt som I många avseenden hade samma fokus som detta projekt, dock omfattande samtliga forskningsområden.

I projektets slutrapport¹⁴ sammanfattas ett antal nyttiga erfarenheter som har legat till grund för centrets fortsatta arbete. Den fokuserar starkt på vikten av samarbete och bred uppslutning kring datafrågorna från samtliga inblandade parter, liksom på vikten av kommunikation och framtagande av gemensamma policies.

2.5 Aktuella verksamheter

2.5.1 Göteborgs universitet

Vid Göteborgs universitet bedrivs bred forskning inom såväl humanistiska som konstnärliga vetenskaper.

Humanistiska fakulteten består av följande institutioner:

Institutionen för filosofi, lingvistik och vetenskapsteori, Institutionen för historiska studier, Institutionen för kulturvetenskaper. Institutionen för litteratur, idéhistoria och religion, Institutionen för språk och litteraturer och Institutionen för svenska språket.

Konstnärliga fakulteten består av:

Filmhögskolan, □Högskolan för fotografi, □Högskolan för Design och Konsthantverk, HDK □Steneby - Institutionen för Konsthantverk och Design, □Högskolan för scen och musik, □Konsthögskolan Valand, □Litterär gestaltning och □Göteborg Organ Art Center, GOArt.

Flera institutioner eller forskargrupper från båda fakulteterna har sedan tidigare kontakt eller påbörjat samarbete med SND.

I projektet intervjuades tre forskare från olika institutioner från Konstnärliga fakulteten samt tre forskare från olika institutioner på Humanistiska fakulteten.

¹¹ <http://edina.ac.uk/>

¹² <http://www.sharegeo.ac.uk/>

¹³ <http://www.disc-uk.org/datashare.html>

¹⁴ <http://ie-repository.jisc.ac.uk/336/1/DataSharefinalreport.pdf>

Möten hölls med dekanerna från båda fakulteterna och ytterligare en forskare intervjuades för en första testintervju.

Universitetet använder e-publiceringssystemet GUPEA¹⁵ för avhandlingar, rapport- och skriftserier, studentuppsatser och även video- och ljudresurser. För GUPEA använder man Open Source- programvaran DSpace. I GUPEA finns sedan tidigare flera exempel finns på avhandlingar och andra forskningspublikationer med tillagd forskningsdata/material.

Universitetet använder en separat publikationsdatabas GUP¹⁶ med obligatorisk registrering av referenser till alla forskningspublikationer av anställda vid Göteborgs universitet. GUP utvecklas på den egenutvecklade plattformen Scigloo, i samarbete med Chalmers Bibliotek.

Från Göteborgs UB deltog Urban Andersson (som projektledare) och Sofia Arvidsson i projektarbetet. I inledningsskedet deltog även Carin Björklund.

2.5.2 Linköpings universitet

Forskningen som kan vara aktuell för detta projekt är finns inom tre institutioner, fördelade på två fakulteter: Institutionen för studier av samhällsutveckling och kultur (ISAK), där Avdelningen för historia ingår, Inst. för kultur och kommunikation (IKK), som inkluderar språk, filosofi, konstvetenskap, litteratutvetenskap, musik och religionsvetenskap, samt Inst. för ekonomisk och industriell utveckling (IEI), som bland annat inkluderar Malmstensskolan.

Den sistnämnda är lokaliserad i Stockholm och har utbildningar inom Möbeldesign, Möbelsnickeri, Möbeltapetsering och Möbelkonservering. Man bedriver i nuläget ingen forskning, men förberedelser för detta pågår.

Ett forskningsprojekt med relevans för detta projekt är ”Konst genom staden” (se vidare punkt 5 nedan).

Linköpings UB har inom projektet haft kontakter med forskare från Historiska institutionen, rörande ett register över medicinhistoriska data. Någon forskare från denna grupp hade sedan tidigare varit i kontakt med SND, angående registrering av metadata i SND:s system.

Man även haft vissa kontakter med forskare inom barnspråk (med en omfattande mängd analogt ljudmaterial), med Avdelningen för historia vid Institutionen för studier av samhällsutveckling och kultur.

Linköpings universitet ingår i DIVA-samarbetet.¹⁷ DIVA är ett system som, i likhet med de flesta övriga befintliga publiceringssystem, inte primärt är byggt för att hantera data. Forskningsdata kommer sannolikt att hanteras i ett annat system, på en annan server.

Från Linköpings UB deltog David Lawrence i projektarbetet.

¹⁵ <http://gupea.ub.gu.se>

¹⁶ <http://gup.ub.gu.se>

¹⁷ <http://www.diva-portal.org>

2.5.3 Lunds universitet

Vid Lunds universitet bedrivs forskning både inom konstnärliga och humanistiska vetenskaper.

I projektet kontaktades inledningsvis forskare vid den konstnärliga fakulteten i Malmö. Dessa kontakter var dock inte fruktbara vad det gällde att få fram ett konkret material att arbeta med där det fanns både data och publikationer så efter en tid försökte man hitta ett sådant vid den humanistiska fakulteten i stället.

Där möttes man av ett positivt intresse och valde tillsammans ut ett projekt som innehöll intressant data och hade genererat ett flertal publikationer.

För beskrivning av arbetet det konkreta exemplet se Språkforskning – Swedia2000 under punkt 5.

Universitetet använder ett egenutvecklat system för publicering och publikationsregistrering, LUP¹⁸.

Från Lunds UB deltog Jörgen Eriksson och Peter Berry i projektarbetet.

2.5.4 Malmö högskola

I Malmö hade man främst kontakter med Inst. för Konst, kultur och kommunikation (K3). Forskningen och det konstnärliga utvecklingsarbetet där är indelat i två teman – ”Design” och ”Kultur och medier”. Kontaktperson var professor Jonas Löwgren.

En av de intervjuade forskarna har i projekt använt metoden Cultural Probes (”dagboksstudier”, se vidare under punkt 5 nedan).

Malmö högskola använder MUEP, ett system baserat på DSpace, som e-publiceringssystem.

Från Malmö högskola deltog Jessica Lindholm och Linda Trygg i projektarbetet.

2.5.5 Svensk nationell datatjänst (SND)

SND har i uppdrag av Vetenskapsrådet att vara en nationell resurs för samordning av existerande och nya forskningsdatabaser inom samhällsvetenskap, humaniora och medicin. SND inrättades den 1 januari 2008 på initiativ av Vetenskapsrådet.

Verksamheten bedrivs vid Göteborgs universitet och finansieras gemensamt av Vetenskapsrådet och Göteborgs universitet. SND skall ”medverka till teknisk utveckling av standarder, metoder och verktyg för dokumentation och tillgängliggörande av dessa data” /... / ”utredning av juridiska problemställningar” /.../ samt ”stödja bevarande, användning och tillgängliggörande av databaser och relaterade forskningsmaterial inom SNDs ansvarsområden, existerande och nya forskningsdatabaser inom samhällsvetenskap, humaniora och socialmedicin/epidemiologi.”¹⁹

¹⁸ <http://lup.lub.lu.se/>

¹⁹ Citat hämtade från SND webbplats <http://www.snd.gu.se/>

Man arbetar primärt med insamling och spridning av information om forskningsdata. Insamling av data sker kontinuerligt och forskare kan lämna data till SND för arkivering och - framför allt - tillgängliggörande.

Stor vikt läggs vid datats återanvändbarhet – det skall vara begripligt, och samtidigt bearbetat (avkodat/avidentifierat).

SND skall även ge stöd och rådgivning åt forskare. Man förfogar över en katalog som är sökbar på nätet. Det är möjligt för forskare att beställa data, men inte Open access. I uppdraget ingår att ta hand om data i olika format.

Man tittar också på andra och kompletterande system, liksom olika möjligheter att dela och tillgängliggöra data i samarbete med andra existerande system (se System nedan).

SND arbetar med tillgängliggörande av forskningsdata inom forskarvärlden. Bearbetad, förädlad forskningsdata.

Man lägger ner ett stort arbete på att avidentifiera data och att kontrollera format för att göra forskningsresultaten möjliga att återanvända. Forskare som vill ta del av data måste kontakta SND med förfrågan och beställning. SNDs insamlade forskningsdata är alltså inte fritt tillgängligt. Det finns planer på att data ska bli tillgängligt via ett inloggningsförfarande.

SND har också i uppdrag att ha en översikt över svenska forskningsprojekt inom sina ämnesområden.

Från SND deltog Iris Alfredsson och Birger Jerlehag i projektarbetet.

3. Genomförande

Den primärt valda metoden är personliga djupintervjuer med forskare inom berörda ämnesområden. Som underlag har använts en av projektet utarbetad grundmall (bilaga 2).

Detta för att kunna presentera jämförbara data från de olika lärosätena. Varje deltagare genomför också en pilotintervju, samt – i förekommande fall – inledande samtal med dekan/prefekt, för att förankra projektet i respektive organisation.

Intervjuerna syftar till att kartlägga till såväl forskarnas syn på och inställning till tillgängliggörande av forskningsdata, som praktisk inventering av vilka typer av data som kan förekomma inom respektive ämnesområde.

Inom projektet intervjuades ett antal forskare om hur de arbetat med forskningsdata i några olika forskningsprojekt, för att samla in mer kunskap och exempel specifikt från humanistisk och konstnärlig forskning. Trots ett litet antal intervjuer har vi fått fram en stor variation i svaren när det gäller flera olika aspekter av forskningsdata. Vi har också fått inblick i den stora variationen mellan olika discipliner inom de båda fälten.

4. Intervjuer, sammanställning

Varje intervju inleddes med en kortare presentation av projektet, dess syfte och metod. I flera fall diskuterade vi utifrån den övergripande skissen av forskningsdataprocessen (bilaga 1). Intervjumall (bilaga 2)

4.1 Forskningsprojekt

I intervjuerna gavs exempel på flera olika slags forskningsprojekt:

- Interaktionsdesignforskning kring ungdomar och samhällsfrågor
- Produktion och användning av audiovisuella medier som hjälpmedel
- Kartläggning av folkhälsa, empirisk studie över attityder och samhälle
- Användning av olika material i rumsskapandet och relation till olika kulturers föreställningar om rum
- Inventering av orglar
- Utveckling av kompositionsverktyg, analys av den egna skapande processen
- Utveckling av metoder för genomförande och utvärdering av konstnärlig forskning inom teaterns område
- Digitalisering av dokumentationsmaterial, bräckliga dokument i stora format
- Forskningsprojekt om kommersiell samtida musik
- Forskning om författares arkiv och författarskap

4.2 Fråga 1 - Insamling och bearbetning av forskningsdata

Under frågorna om insamling av material gavs en mängd exempel på olika slags forskningsdata; anteckningar, foton, videofilmer, intervjumaterial, notmaterial, skisser, prototyper, data med visuellt fokus, bild, film, bildspel, utställning, ljudfiler, föreställning, egenproducerade texter, kartor, ritningar och kataloger. Även datainsamling utifrån en modell där det ingick material där informanterna svarade genom att arbeta med vykort, fotografering med engångskameror, anteckningsböcker. Några angav att deras forskningsdata var litteraturbaserad: artiklar, material från andra forskare, pdf:er eller utskrivna papperskopior av artiklar.

En forskare anger kartor, ritningar som forskningsdata. Denna forskare har åkt runt och besökt museer och hämtat in material. Det finns inte mycket digitalt, dock en del digitala kataloger som forskaren använder. Forskaren upprättar egna förteckningar över sitt material samt kommentarer.

Primärdata är skriftligt tryckt material från lärosätets bibliotek. Andra typiska är ljudfiler och notmaterial, och en viktig källa är Kungliga bibliotekets (KB) audiovisuella avdelning. Forskaren kontakter KB:s audiovisuella avdelning, de digitaliserar på efterfrågan och forskaren får sedan tillgång till filerna. Forskaren har ofta arbetskopior under tiden för projektet.

En forskare spelar själv in mycket ljudfiler, liksom fler och fler videofilmer - i projektet behövs videoinspelningar eftersom verktygen de arbetar med kräver mänsklig interaktion och det tillför dessutom något för publiken att se på dem som spelar och vad som samtidigt händer i musiken. Forskaren analyserar även sin egen skapelseprocess och skapar en metod utifrån det. Forskaren analyserar också hur de

arbetar med de nya verktyg som de tar fram för att komponera musik och för att skapa nya ljud.

En av forskarna skapar mycket av forskningsmaterialet under projektets gång; två föreställningar, ett manus till en pjäs, skådespelarnas loggböcker (utarbetad metod), beskrivning av forskningen i programblad, artiklar, kortare texter, två dvd:er varav en publicerad fritt tillgängligt.

En forskare producerar ofta eget unikt nytt material i form av texter, dikter, romaner. Materialet bearbetas många gånger under projektets gång. Texterna är ofta författarens egna så upphovsrätten tillhör författaren. Forskaren i detta fall vill inte ge ut texter i universitetets namn, därför har en ideell förening skapats där de hittills gett ut nio böcker och en del nätbaserade texter. Även av skälet att de då kan erhålla kulturstöd.

Ett exempel fanns på en analys som forskaren bara skickat i form av e-post som innehöll en sammanställning samt ett antal webbplatser som forskaren hänvisade till. Forskaren anger att den eventuellt finnas kvar i mailen annars finns det inte dokumenterat. Någon forskare angav att man arbetar en hel del med andras material där det finns mycket regler för sekretess och annat ex intervjumaterial, patientjournaler, enkätstudier. Enkät svar anges av ytterligare en forskare och i ett projekt var det även ljudinspelningar på DAT-band

Forskaren åker runt till arkiv och letar i dem på plats, väldigt lite finns digitalt. Ger exempel på hur processen att hitta material; att hitta kvinnors självbiografier kan vara svårt t ex om de levt ihop med en känd man: "Mitt liv med..." och material är klassificerat på mannens namn. En del kan också vara svårt att avgöra vad det är för material, forskaren har ibland utelämnat material eftersom det varit för osäkert för att ta med. Forskaren antecknar på datorn och i kollegieblock och inte strukturerat. Anteckningar bara för sig själv, ingen strukturerad dokumentation, för att göra detta skulle det krävas att det t ex hade betydelse för meritering alternativt fanns med i anslagen.

En forskare menar att om man skall kunna tillgängliggöra forskningsdata för andra måste detta in som ett krav i ett tidigt skede så att man kan anpassa insamlandet av material efter det. Skulle ta väldigt mycket tid men vara användbart eftersom en del arkiv är "otroligt jobbiga" att hitta något i. Tycker det finns en problematik också att andra kan ifrågasätta ens material (antagligen hellre än att validera det), alltså att materialet kan komma att användas emot forskaren.

En kommentar som kom upp var att forskningsdata, i synnerhet inom konstnärlig forskning, ofta i sig utgör ett konstnärligt arbete (bild, manuskript, musikstycke etc.) och att detta kan vara svårt att avskilja från den faktiska publikationen. Här finns möjligen en viktig skillnad gentemot andra forskningsområden, där forskningsdata har en mer naturligt stödjande roll i forskningsprocessen.

4.3 Fråga 2 - Lagringsmedium

Forskarna anger att deras forskningsdata lagras på bl a papper, VHS-kassett, CD-skivor, extern hårddisk, egen databas, webbfiler, inscannade bilder, fotoformat i

digital kamera, institutionens server, egen hårddisk hemma,, "lagring i kartong på institutionen" och "i huvudet". Ofta lagrar forskarna ostrukturerat och utan annotationer. Flera forskare anger att de sparar forskningsdata både på institutionens server och på sin privata dator hemma och dessutom tar back-up på båda ställen.

En forskare lagrar artiklar och anteckningar på sin dator och på institutionens server och kommenterar att "man vet att andra forskare sparar ofta materialet". Forskaren lägger ut en del färdiga resultat på academia.edu (parallellpublicering), lagrar filer på institutionens server och har back-up på personlig dator hemma.

En forskare sparar på CD-skivor och externa hårddiskar. Forskaren tycker inte det är bra lösning och har tagit upp detta på fakulteten. De har 10 GB lagringsutrymme på servern per person vilket forskaren tycker är alldeles för lite. Forskaren tar egen back-up. Forskaren är mycket intresserad av att frågan om långtidslagring blir löst. Forskaren vill kunna "lyfta bort" färdigt material från servern. En forskare arbetar med en egenutvecklad databas som finns på en server på fakulteten. Forskaren menar att säker lagring ändå inte är garanterat: "säker lagring är som ett stort svart hål".

En forskare sparar filerna i mappar på jobbet och verkar kunna tyda dem själv, ibland med hjälp av kalendern (stämmer av med datum vilken inspelning som hör till vilken konsert) men menar att de generellt "har mycket i huvudet". Lagrar både på jobbet och hemma, har bra back-up på båda ställen.

En forskare sparar sina arbetskopior och är noga med att göra säkerhetskopior. Forskaren tycker dock att materialet är oftast hämtat från arkiv som där tar hand om lagringen över tid. En forskare anger dator, block, i huvudet. Forskaren kan vid publicering i princip slänga anteckningarna, då dessa inte längre behövs.

4.4 Fråga 3 - Publicering av resultaten från forskningsprojekten

Resultaten från forskningsprojekten publiceras/tillgängliggörs t ex som artiklar, böcker, bokkapitel, rapporter, konferensbidrag, avhandling, utställningar, böcker (ex även där det ingår ett urval av foton), föreställningar, kompositioner, ljudinspelningar, videoinspelningar, konserter, installationer och multimedia.

En forskare publicerar artiklar, kapitel, konferensbidrag, kompositioner, ljudinspelningar, videoinspelningar, har konserter, installationer etc. Ibland med illustrerande ljudfiler, cd eller dvd som följer med i böcker. Forskaren tror att det är avgiftsbelagda tidskrifter som de använder, men känner till att finns angivet i avtalen att man får lägga ut en kopia på sin egen webbsida och gör därför oftast detta. Forskaren tycker det blir mer problematiskt när det gäller inspelningar eftersom forskaren är STIM-ansluten. De behöver t ex ofta ansöka om licens för att lägga ut klipp på webbsidor.

En forskare har lagt länk till en presentation från konferens med innehåll som mer varit på ett idéstadie från en bibliografisk post i lärosätets publikationsdatabas. Forskaren har lagt ut mycket på academia.edu fritt tillgängligt, även på institutionens server. En forskare publicerar artiklar, föreställningar och inspelade manifestationer av projektet (DVD), oftast utan tilläggsmaterial. Artiklarna publiceras fritt tillgängligt i bl a nättidskrifter.

En forskare har producerat en slutrapport till uppdragsgivaren som varken finns tillgängliggjord i tryckt form eller digital. Beroende på projekt publicerar de resultat i tidskrifter. En forskare publicerar främst böcker. Forskaren har provat att lägga ut en text med tillhörande bilder på en webbsida men fick inte så mycket kommentarer på detta som forskaren trodde. En forskare publicerar monografier, forskaren menar att det finns ett stort undermaterial till en monografi, allt kommer inte med- ibland publiceras dock tilläggsmaterial i bilagor.

Forskaren publicerar sig i tidskrifter, bokkapitel och monografier. Det finns en peer reviewed tidskrift för forskningsområdet i Sverige. Forskaren menar att den borde ha förnyat sig för länge sedan och bytt språk. Flera fritt tillgängliga tidskrifter har kommit inom forskarens forskningsområde de senaste 2-3 åren vilket forskaren tycker är positivt för ämnesområdet. De som forskar om kommersiell samtida musik kan inte lägga additional files av upphovsrättsliga skäl och forskaren menar att "alla vet att man ändå kan lyssna på det på Youtube eller Spotify". Forskaren anger att frågan diskuteras på ämnesspecifika e-postlistor.

4.5 Fråga 4 - Tillgängliggörande av forskningsdata

De flesta forskarna tycker att fri tillgång till forskning och underliggande forskningsdata är viktigt och är i allmänhet mycket positiva. Flera olika skäl anges som orsaker till detta, men olika invändningar tas också upp. Flera har lagt publikation eller annan manifestation av forskning, samt ett urval av forskningsdata på institutionens eller egen webbsida, alternativt i annat publiceringssystem på internet.

Tre forskare med stort geografiskt avstånd till Göteborg (där SND är lokaliserade) har inte haft kontakt med dem. Flera forskare från Göteborgs universitet har dock haft kontakt med SND och även redan inlett samarbete i vissa fall.

När det gäller att tillgängliggöra forskningsdata så tycker en forskare att visst material, t ex filmade diskussioner och intervjuer kan passa att tillgängliggöra via SND för andra forskare, i de fall där materialet är mer känsligt. Forskaren känner väl till lärosätets öppna arkiv men tycker det är oklara rutiner kring hur man publicerar där, forskaren använder istället academia.edu för att det är så enkelt.

En forskare tycker det är mycket intressant och viktigt att tillgängliggöra forskningsdata men är dock rädd om sitt material, och ser därför en ytterligare fördel med tillgängliggörande hos SND är att man som primärforskare blir personligt kontaktad. Forskaren tycker också det kan vara viktigt för forskarnas kontaktnät.

En forskare är väldigt positivt till tillgängliggörande men tror mer på tillgängliggörande via SND eftersom "internet är alldeles för spretigt och okontrollerat".

Forskaren menar att det är lätt att omedvetet använda saker på fel sätt.

Grundsynen är att forskningsmaterial skall vara tillgängligt för andra och att det är bra om det är bearbetat så att någon kan begripa det. Forskaren känner inte direkt till lärosätets öppna arkiv men har däremot lagt in det mesta i lärosätets publikationsdatabas. Forskaren känner igen namnet SND men vet inte så mycket mer.

Det finns inget ämnesarkiv inom forskningsområdet, använder t ex istället Myspace, YouTube och Vimeo för att sprida verk.

Ytterligare en forskare tycker att för tillgängliggörandet till en bredare allmänhet är t.ex YouTube intressant.

Forskaren tycker SND är viktigt och bra och tycker att visst material kan passa här – exempelvis för sådant som har etiska aspekter. Forskaren känner till lärosätets öppna arkiv men menar att den seniora forskningen saknas, avhandlingarna finns men inte annan forskning inte heller t ex mastersstudenters konstnärliga projekt.

Forskaren tycker att materialet skall finnas fritt tillgängligt för alla och kommenterar inte så mycket om ”graderna” av tillgänglighet. Menar att det som ingår i en nationell inventering självklart skall vara fritt tillgängligt, men tycker att det finns annat material som är forskarens eget. Tycker att tillgängliggörande ger en demokratisering av forskningen och att det dessutom sparar tid och resurser eftersom forskarna slipper göra om varandras jobb, dvs att åka runt till olika museum och arkiv och samla in samma material.

Forskaren samarbetar redan med SND i ett projekt. Forskaren menar att fler och fler kataloger och material från museum/arkiv digitaliseras och samlas ihop i databaser.

Forskaren använder såväl KB:s ljud & bildarkiv och fritt tillgängliga tjänster som Youtube och Spotify, som motsvarande ämnesarkiv. Forskaren känner till SND. Forskaren ger exempel på transkribering av samtida hit-låt: de upphovsrättsliga frågorna för att tillgängliggöra något sådant skulle vara alldeles för tidskrävande och oklart vad som gällde. Forskaren gav exempel på kollegor som försökt för att de exempelvis har velat lägga ut något ljudmaterial som komplement till något resultat, men menar att de flesta oftast ger upp.

“Det vore förstås förträffligt men är det värt besväret att kolla upp alla upphovsrättsliga aspekter? Vad blir kvar om man väger intresse, integritet, upphovsrätt osv?”.

Forskaren vill att forskningsdata skall vara tillgänglig, men med begränsad åtkomst för andra forskare (SND) - och tror att det är denna lösning som kommer att fungera i praktiken. En annan tycker att detta är en sämre lösning, eftersom man då väljer ut vilka som blir användare och materialet då ej blir tillgängligt för alla. Den sistnämnda forskaren är mycket positivt inställd till fritt tillgängliggörande för alla och menar att detta kan skapa synergieffekter. Forskaren tror att det skulle utveckla humaniora och menar att man inom detta forskningsområde ibland sitter fast med samma arbetsmetoder som man använt sedan 1800-talet - “det är bra om de måste tvingas till dialog”. Menar att även de unga forskarna ibland är mycket hemliga med sitt material. Farhågor med fritt tillgängligt för alla kan vara ”elakheter inom akademien”, dvs att materialet kan användas emot forskaren på ett negativt sätt.

En forskare efterlyser ett ämnesarkiv (men känner själv inte till något sådant arkiv), en forskare anger att det finns inget ämnesarkiv inom ämnet, men finns inom närliggande ämnen. En annan forskare nämner att det eventuellt finns ett arkiv inom närliggande ämne. En forskare anger att det inte finns något specifikt ämnesarkiv- forskaren hittar annan forskning genom personligt nätverk, webben och genom att ”botanisera i bibliotek”.

En forskare nämner etiska frågor, loggböckerna i forskarens projekt är mycket personliga och måste också vara på det sättet- de medverkande måste kunna skriva fritt, loggböckerna ser forskaren som ett mycket viktigt men samtidigt problematiskt material.

4.6 Fråga 5 - Återanvändning av forskningsdata

Några forskare är tveksamma till om deras material är meningsfullt att återanvända men flertalet tycker att det är intressant för andra forskare, även för forskare inom andra ämnesområden. Flertalet anger att det krävs mycket bearbetning av materialet för att detta skall kunna återanvändas. Flera tar upp olika etiska aspekter och även sådana som omfattar avidentifiering.

En forskare tar upp exempel med VHS-band med inspelningar med fokusgrupper för ett projekt. Detta kan också vara intressant för andra forskare inom ett helt annat forskningsområde. Forskare tyckte att bearbetning inför att tillgängliggöra materialet skulle gå väldigt fort för just detta projekt, krävs i princip bara digitalisering. Invänder dock att det fanns annat material som forskaren skulle dra sig för att lägga tid på att tillgängliggöra. Bearbetning tar generellt väldigt mycket tid och forskarna har I allmänhet inte tid med detta.

En forskare menar att forskningsmaterialet, i huvudsak bestående av fotografier på byggnader, skulle vara mycket användbart för någon annan, då det kan vara svårt att hitta vissa byggnader. Kommenterar dock att man ju samtidigt oftast inte skriver om det man själv inte sett på plats. Forskaren tycker dock ändå det skulle kunna vara användbart.

En forskare tycker inte att materialet är så användbart i just det projekt som forskaren utgår ifrån i diskussionen, men menar att det finns situationer när det skulle kunna vara det. Exempelvis om kyrkan med orgeln som det forskats om skulle brinna ner. I det fallet skulle även forskarens oredigerade inspelningar kunna visa sig vara mycket värdefulla. Det är en bedömningsfråga hur lång tid som skulle behövas för bearbetning och detta beror på hur man skall använda materialet. Forskaren drar en parallell till katalogisering - man kan begränsa till få uppgifter eller vara mer utförlig. Forskaren skulle behöva kommentera och dokumentera något, men inte mycket i det aktuella forskningsprojektet.

Forskningsmaterialet skulle vara användbart och forskaren är positiv till att det skulle användas "om någon frågar får de gärna använda materialet". Skulle dock krävas mycket bearbetning och dokumentation för att någon annan skulle kunna använda materialet. Det skulle krävas "enormt arbete" enligt forskaren.

Materialet skulle vara mycket användbart för andra. Det skulle krävas ganska mycket bearbetning men forskaren menar att denne ändå tänkt göra detta och är mycket positiv till det, har dock inte har tid till detta nu men däremot under nästkommande termin.

Forskaren tycker att materialet mycket väl kan återanvändas, forskaren har flera gånger låtit doktorander ta kopior av materialet, en del material behöver dock digitaliseras.

En forskare kommenterar att man förr ofta skrev så att det var nödvändigt att även ta del av texten transkription, men det gör man inte idag. Det händer att man skulle vilja ha med det - liksom även notexempel - men att man oftast avstår av upphovsrättsliga skäl.

Materialet skulle vara användbart för andra - ”ofta vill man ju att andra skall börja forska inom ens område”. Det skulle dock ta väldigt mycket tid i anspråk att bearbeta materialet.

4.7 Fråga 6 - Spridning genom länkning mellan publikation och forskningsdata

De flesta forskarna är positivt inställda till länkning mellan publikation och tillhörande data. Två forskare har själva på prov skapat en webbsida med länkning mellan publikation och data. En forskare anger att det är vanligt med additional files med skisser och processbeskrivning till publikationer inom fältet.

En forskare förordar SND:s lösning för det större oredigerade materialet, forskaren tror att det slarvas mycket med copyright på material som man hittat på Internet och tycker inte det är bra att material rycks ur sin kontext. Ett mindre mer kontrollerat urval vore däremot i sin ordning att lägga ut.

En forskare diskuterar om det skulle vara en länk till själva materialet direkt åtkomligt, isåfall skulle bearbetning behövas om det skulle vara begripligt för någon annan. Å andra sidan om det skulle vara en referens till att materialet fanns på SND skulle ändå bearbetning behövas för att lägga det hos SND.

En forskare tycker att det skulle vara bra men oftast inte möjligt då det är kommersiella ljudfiler forskaren oftast utgår från och man kan inte länka till Youtube - “som är en upphovsrättslig gråzon” - eller Spotify.

4.8 Fråga 7 - Hjälp

En forskare ger ett konkret exempel: forskaren vill att det skall finnas en tjänst som tar hand om upphovsrättsliga frågor och frågor om säker lagring och bevarande. Det spelar mindre roll vart tjänsten finns; hos KB, SND eller lärosätets bibliotek. Forskaren tycker inte det är bra att alla sitter och bygger upp egna databaser och uppfinner hjulet gång på gång. Forskaren betonar att SND kommer att spela en viktig roll i framtiden.

Behovet av tekniskt stöd i olika delar av processen betonas också.

4.9 Fråga 8 - Deltagande i projektet

De flesta av forskarna är positiva till att delta med forskningsmaterial i projektet. Några anger dock att materialet behöver bearbetas, digitaliseras eller att de inte har tid för deltagande vid tillfället men gärna längre fram.

En forskare tycker att viljan att dela med sig av forskningsdata, utkast och idéer kan bero på hur etablerad man är som forskare. En del etablerade forskare har mer ”råd” att dela med sig. Om det är ett nydanande resultat vill forskaren nog hålla det för sig själv tills dess att det publicerats, eftersom det kommer att påverka forskarens arbetsvillkor. Forskaren lägger ändå själv ut en del idéer eller utkast för en mindre grupp via e-post, ibland genom presentationer på konferenser, för att få hjälp och råd från kollegor.

En forskare vill tillgängliggöra foton från ett forskningsprojekt och vill att dessa skall finnas fritt tillgängligt men inte nedladdningsbara. Forskaren vill helst inte ens att en lågupplöst JPEG skall kunna gå att fritt ladda ner. Samtidigt vill forskaren inte lägga just det materialet hos SND ”då blir det ju inte fritt tillgängligt”. Forskaren undrar om det kan ordnas så att filerna inte går att spara ner. Forskaren har också en text som beskriver en utställning och ett bildspel som visades på utställningen. Detta ihop med ett urval av foton, t ex 4 st kan forskaren tänka sig att lägga helt fritt tillgängligt

En forskare deltar genom att skicka sin avhandling i fulltext samt de ljudfiler som tillhör den för publicering i lärosätets öppna arkiv. Forskaren deltar redan i angränsade projekt genom påbörjat samarbete med SND om ett större antal bilder de digitaliserat som saknar metadata. De har digitaliserat materialet men ännu inte bedrivit forskning på det. Om det tillgängliggörs på SND slipper de lagrings- och hanteringsfrågan samtidigt som de kan starta sin egen forskning på materialet och andra kan samtidigt starta forskning ur andra aspekter.

Två forskare kan tänka sig att delta i projektet, förslag är med material från ett projekt, har dock inte tid just nu och vill återkomma i mitten av sommaren eller senare i höst. En av dem skickar en förteckning över sitt material för senare kontakt.

En forskare föreslår först sin egen katalog eftersom forskaren kommer att skriva en bok under året och då skulle det vara bra om den blev digitaliserad och tillgängliggjord. Forskaren föreslår sedan kataloger på CD-rom som då redan är digitala, som finns som bilagor till tryckta avhandlingar. På det sättet skulle projektet få in något material snabbt. Forskaren tycker att vi skall kontakta de arkeologiska institutionerna som finns i Sverige och be om deras kataloger för att lägga ut.

En forskare vill inte delta alls, har inget material som skulle passa för det. Menar att passande material till SND skulle kanske vara intervjuer och passande material för Open Access--publicering skulle vara sådant forskaren själv var upphovsman till. Exempel på det kunde vara en doktorand som spelat Beethoven på piano och gjort inspelningar på detta. Upphovsrätten till kompositören har då gått ut och doktoranden är själv upphovsman till både framförande och inspelning.

En forskare är positiv till deltagande, kan tänka sig att delta med en bibliografi från ett appendix i sin avhandling, 15 sidor som dock behöver digitaliseras.

5. Typer av data – problem, utmaningar och lärdomar

Forskningsdata inom humanistiska och konstnärliga vetenskaper kan, av naturliga skäl, förekomma i en mängd olika format. Dessa format kan vara mer eller mindre applikationsspecifika, liksom mer eller mindre proprietära. Båda dessa saker påverkar naturligtvis datats tillgänglighet och återanvändningsgrad, två centrala begrepp om forskningsdata skall tillgängliggöras open access.

I många fall är det sparade materialet i analog form – anteckningar, foton, ljud- och filminspelningar – och måste först digitaliseras. I andra fall finns materialet redan i digital form, men måste ändå bearbetas, alternativt konverteras, för att kunna lagras säkert och användas på ett meningsfullt sätt.

Graden av bearbetning kontra rådata kan vara ett viktigt ställningstagande i varje enskilt fall.

Vidare så måste alla former av forskningsdata förses med beskrivande metadata och dokumentation för att göras sökbart, begripligt och återanvändningsbart.

I projektet studerades ett antal olika fall i praktiken. Av dessa kan tre nämnas särskilt:

Språkforskning - Swedia 2000.

Projektet Swedia 2000 har samlat in talprover från över 100 svenska dialekter. För varje dialekt har man spelat in minst 12 olika talare – äldre kvinnor, äldre män, yngre kvinnor, yngre män. Projektet är avslutat.

Talproverna finns som okomprimerade vaw-filer och som mp3-filer. Till varje talprov hör också textfiler med en transkription av talet och en översättning till normalsvenska.

Datat finns idag på en server i Humanistlaboratoriet i Lund. En liten del finns fritt tillgängligt genom Inst. för lingvistik vid Göteborgs universitet²⁰, som även listar projektets publikationer.²¹

Många av dessa länkar till fritt tillgänglig fulltext. Av de totalt 53 listade publikationerna kunde 36 läggas ut som fritt tillgängliga i Lunds institutionella arkiv.

En fanns fritt tillgänglig i Umeås arkiv och 3 las upp i Göteborgs arkiv.

Forskarna vid Humanistlaboratoriet var positiva att göra datat tillgängligt och la upp det för nerladdning av SND efter att ha plockat bort bilder som fanns på informanterna i projektet. SND fick också en lista över publikationerna med länkar till de institutionella arkiven.

Vad som återstår att göra för att skapa kopplingen mellan projekt och publikationer är att SND katalogiserar datat och gör det tillgängligt samt listar publikationerna med länkar till de institutionella arkiven. För dessa återstår att länka tillbaka till SNDs post från de enskilda publikationsposterna.

²⁰ <http://swedia.ling.gu.se>

²¹ http://www.ling.gu.se/~anders/SWEDIA/publ_sv.html

Stort tack till Susanne Schötz och Nils Holmberg vid Humanistlaboratoriet i Lund för ett gott och positivt samarbete.

Cultural Probes

Denna metod har använts av en forskare vid Malmö högskola som vi hade kontakt med genom projektet.

Cultural probes, även benämnt “diary notes”²², är en metod som syftar till att samla in kunskap om en grupp människor, deras interaktion och den kultur som de verkar i, genom att förse dessa personer med en uppsättning verktyg för att dokumentera sin vardag. I det här fallet utgjordes dessa av engångskameror, anteckningsmaterial, vykort och fiktiva kartor och forskningsdata i sammanhanget utgjordes av en låda där detta råmaterial helt enkelt hade samlats.

För att lagra och tillgängliggöra detta data på ett tillfredsställande sätt så skulle man behöva digitalisera materialet, såväl anteckningar och vykort, som resultatet från de engångskameror som använts. Spara detta i vedertagna format, samt beskriva materialet med såväl metadata, som ingående dokumentation av innehållet och metoden för att göra det användbart för kommande forskning. Förmodligen skulle det i det här fallet också vara aktuellt att, av integritetsskäl, aidentifiera delar av materialet.

Konst genom staden (Art through city space)

Ett forskningsprojekt som pågick 2005-2008 och som syftade till att utveckla konstnärlig forskning som på olika sätt berörde den staden som rum. Inst. för studier av samhällsutveckling och kultur vid Linköpings universitet deltog i projektet, som leddes av Stockholms universitet och finansierades genom Vetenskapsrådet anslag för konstnärlig forskning och utveckling.

Projektet har relevans i detta sammanhang, då resultatet – rapporttext, intuitivt sammanlänkad med bearbetade data - finns fritt publicerat online.²³ Detta utgör ett mycket gott exempel på hur publikation och ingående forskningsdata kan presenteras.

Några lärdomar från SND

Projektet har i mångt och mycket varit en lärotid för SND vad gäller forskningsdata från humaniora. Det har visat på den diversitet som förekommer, både vad gäller dataobjekt, format och användningssätt. Projektet har i hög grad påverkat den riktning SND nu arbetar efter, och bör kunna märkas praktiskt under hösten, då det under våren anställts en forskningskoordinator med inriktning mot humaniora och konstvetenskap.

Under projektiden har SNDs arbetssätt och rutiner ändrats avsevärt med anledning av juridiska skäl. Till skillnad från publikationer ägs forskningsdata inte av den

²² "What is Cultural probe?". - <http://www.infodesign.com.au/ftp/CulturalProbes.pdf>

²³ <http://www.ep.liu.se/ea/temaq/2010/001/>

enskilde forskaren utan av universitetet/myndigheten där denne är anställd. En nöd att knäcka har varit under vilka omständigheter det går att överföra data från ett universitet, till SND, som ju är del av ett annat universitet.

För att data ska kunna deponeras hos SND ska följande kriterier vara uppfyllda;

- Data går att använda för ny forskning,
- Data är dokumenterade så att de är fullt begripliga för tredje part,
- Data finns i format som möjliggör användning

I projektet fick vi kontakt med ett forskningsmaterial som uppfyllde ovanstående kriterier, SweDia. SweDia 2000 genomfördes under åren 1998-2003 och resulterade i en stor datasamling bestående av ljud, text och bild. Ljuddelen består av redigerade och märkta talprover. Textdelen består av utskrifter och transkriptioner av talmaterialet. Bilddelen utgörs dels av kartor och dels av natur- och kulturgeografiskt bildmaterial. Dessa är organiserade efter landsdel, landskap, (25 svenska + 4 finlandssvenska). Varje landskap består sen av två till fem orter, där fyra individer per ort, en äldre manlig och kvinnlig, samt en yngre manlig och kvinnlig utgör själva objekten i materialet. Hur man för långtidsförvaring skall strukturera detta datamaterial enligt gängse dataarkivmodell är något som ännu inte är helt klarlagt. En kopia av datamaterialet om 250 GB överfördes till SND på hösten 2010. Då denna datamängd är för stor för att i rymmas i SNDs nuvarande datalager, ska den långtidslagras hos SNIC, Swestore, som kommer att utgöra SNDs framtida datalager. Här finns dock inte några färdiga rutiner för struktur och märkning framtagna, så inga permanenta länkar har kunnat anges.

Utöver datamaterialet finns det ett stort antal publikationer, 53 st, varav flertalet i fulltext i LUP, Lunds universitet, samt projektets webbsida som presenterar och tillgängliggör datamaterialet, dock ej fullständigt.

Redan här uppstår frågan om vad som ska länkas till vad, skall publikationerna länka till projektets hemsida eller till arkivkopian hos SND? Ett annat problem är äganderätten, tre universitet är delaktiga i datainsamlingen och en del av projektledarna är pensionerade eller avlidna.

SND testar och utvärderar en programvara (DataVerse) som låter forskaren själv ladda upp, administrera och dela data, lite av samma tanke som finns i dagens OA-system för publikationer. Enligt intervjuaren från projektet är detta något som efterfrågas. Med en sådan lösning får forskningsdata en beständig identifierare vid uppladdning som forskaren sen kan sätta in i sin(a) publikationer. Det innebär även att alla data, även de som inte uppfyller punkterna ovan, kan deponeras, och frågan om äganderätt neutraliseras.

SND håller redan nu på med anpassa sitt publikationsbibliotek för "harvesting" av SwePub. Framöver kommer även datamaterialen vara möjliga att visas upp via SwePub.

Sedan årsskiftet är SND medlem av DataCite, en organisation som tillgängliggör beständiga identifierare i form av DOI:ar (Digital Object Identifier) till dataset. Utöver den beständiga länken lyfts även den bibliografiska citeringen av data upp till ett högre plan. En DOI åtföljs av en hel del metadata som laddas upp till ett nät av datorer som håller ordning på de enskilda objektens status och ursprung. SND tänker givetvis låta andra institutioner som lagrar eller hanterar forskningsdata

att få använda sig av denna tjänst. Idag sitter en prislapp på varje DOI, men förhoppningen är att VR ska betrakta detta som en del av forskningsinfrastrukturen och stå för medlemskostnaden i DataCite.

Övrigt

Givetvis förekommer även inom den humanistiska och konstnärliga forskning forskningsdata som till format och innehåll, liksom nyttjande, är närliggande de som vanligen förekommer inom de samhällsvetenskaperna och som redan idag hanteras av institutioner som Svensk nationell datatjänst. Av det vi har kommit i kontakt med genom projektet kan nämnas olika former av textbaserad dokumentation, såsom loggböcker (teatervetenskap), enkätdata, kataloger och geospatiella data.

6. Systemval

Vi ser att det primärt finns två olika möjligheter att lagra forskningsdata:

- 1) Dataobjekt lagras och tillgängliggörs tillsammans med publikationen, i ett och samma system, som en bilaga, på motsvarande sätt som idag görs med exempelvis fulltextartiklar.
Exempel på detta, även inom de ämnesområden som detta projekt omfattar, finns i berörda system.²⁴

Då i stort sett alla institutioner idag har tillgång till system som i någon mån är förberedda för detta – det vill säga system som i någon mån kan hantera data av olika typer och format – och där kunskap och rutiner redan existerar, så är detta antagligen den mest framkomliga vägen för de institutioner som vill börja hantera forskningsdata rutinmässigt.

Den stora nackdelen är att dagens befintliga system inte är särskilt anpassade till de särskilda krav som hanteringen av forskningsdata ställer. I första hand gäller detta beskrivningen av data, då data- och metadatamodeller i befintliga system fokuserar främst på beskrivningen av publikationen som sådan och inte bifogade filer. För att till fullo kunna hantera forskningsdata så krävs förmodligen mer utförliga beskrivningsmodeller – exempelvis format, typ, version, men även andra beskrivningsfält - av själva dataobjekten. Med andra ord en modell där dataobjekt hanteras parallellt med publikationen och inte som bilaga.

Vidare krävs förmodligen bättre och i hög grad anpassningsbara sök- och presentationsgränssnitt för att göra dataobjekten fullt tillgängliga – inte minst gäller detta objekt inom konstvetenskaperna - liksom standarder som stödjer både export och höstning av komplexa objekt via OAI – t ex OAI-ORE. De befintliga systemen har inte heller alltid lika utvecklad funktionalitet för att hantera åtkomstkontroll, som man har när det gäller behörighetskontroll vid registrering.

²⁴ Se t ex <http://gupea.ub.gu.se/handle/2077/15924> respektive <http://gupea.ub.gu.se/handle/2077/24103>

Vi har också sett att hanteringen av dataobjekt i vissa avseenden skiljer sig markant från hantering och beskrivning av publikationer, och att erfarenhet, kunskaper och rutiner för det förstnämnda kan saknas inom de organisationer som idag primärt arbetar med det sistnämnda.

- 2) Dataobjekt lagras i system som är särskilt konstruerade för detta ändamål, med avseende på såväl datalagring, bevarande och export/import-funktioner, som på beskrivningsmodeller.

Olika typer av data kan lagras i olika system och på olika plattformar, beroende på vilka specifika krav som ställs och man är inte nödvändigtvis begränsad av geografiska och organisatoriska faktorer.

I en sådan lösning förutsätts att dataobjekten är försedda med någon form av unika och persistenta identifierare, exempelvis Data Object Identifiers (DOI)²⁵, och att detta möjliggör sömlös länkning mellan data och publikation. I det ideala fallet tillhandahåller systemet något form av API som kan användas för att dela data på ett för slutanvändaren tillfredsställande sätt. Ett snabbt växande antal exempel på detta hittar man (företrädesvis) inom den naturvetenskapliga forskningspubliceringen.²⁶

All kontroll av versioner, åtkomst och övriga rättighetsaspekter hanteras av systemet som ansvarar för lagringen, och informationen om detta tillhandahålls för publikationssystemet (och andra system) på ett standardiserat sätt.

Vi förutsätter att ett sådant system förvaltas i en organisation där kompetens och erfarenhet av, liksom rutiner för, att hantera forskningsdata redan i någon mån existerar.

Ett annat verktyg för hantering (lagring, publicering och delning) av forskningsdata är plattformen DataVerse.²⁷

Detta är en nätverksbaserad lösning där forskningsdata, tillsammans med beskrivande metadata och länkar till publikationer, samlas i lokala noder, s.k. dataverses, ofta kopplade till enskilda institutioner eller forskningsprojekt. Detta hålls samman i ett centralt DataVerse network.

De enskilda objekten identifieras med s.k. handles²⁸, ett system som liknar och är kompatibelt med ovan nämnda DOI.

DataVerse är baserat på öppen källkod och utvecklas primärt vid Harvard university. Det används bland annat av The Institute for Quantitative Social Science at Harvard University (IQSS DataVerse Network)²⁹

²⁵ "DOI for Datasets", <http://datacite.org/whatisdoi.html>

²⁶ <http://dx.doi.org/10.1016/j.margeo.2004.03.017> refererar till en vetenskaplig artikel, publicerad i tidskriften Marine Geology. Artikelreferensen (i ScienceDirect) länkar i sin tur in forskningsdata från PANGAEA, med DOI doi:10.1594/PANGAEA.727522

²⁷ <http://thedata.org>

²⁸ <http://www.handle.net>

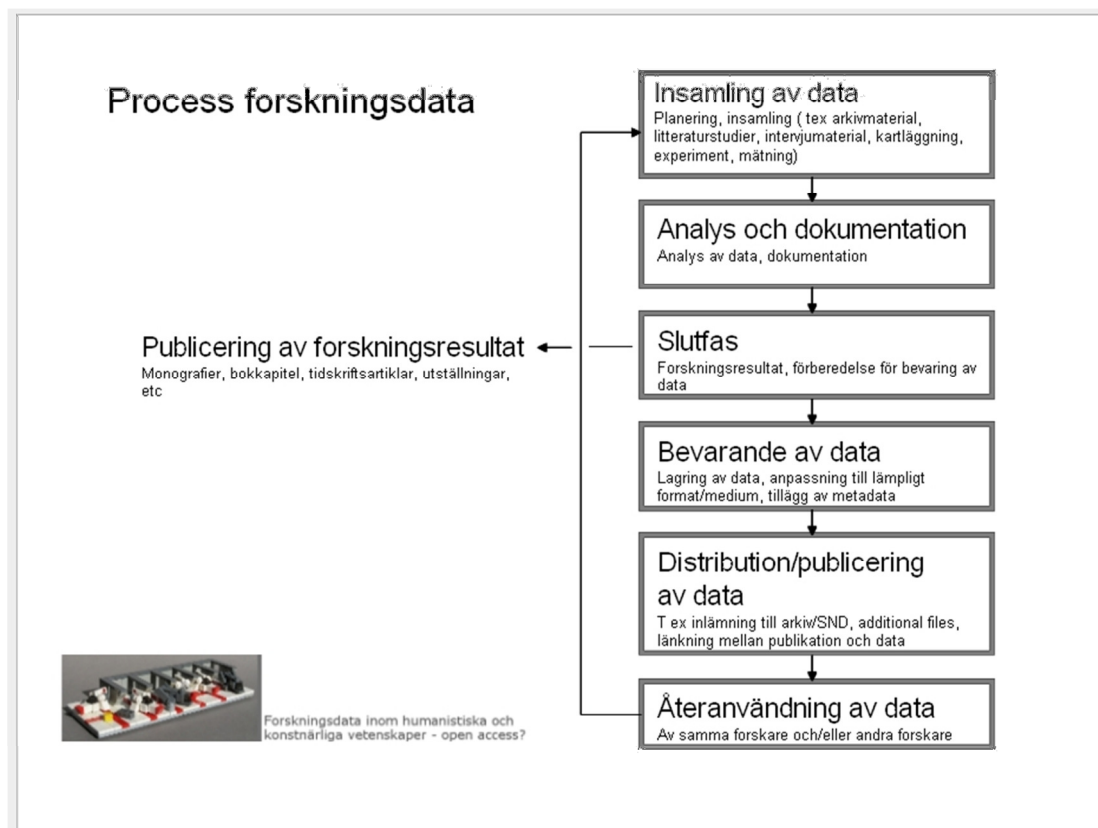
²⁹ <http://dvn.iq.harvard.edu/dvn/>

Svensk Nationell Datatjänst är medlem av DataCite och har en DataVerse server för test av detta system.

Naturligtvis så utesluter inte den ena möjligheten per definition den andra. Från vad som har framgått i såväl intervjuerna och kartläggningen av befintliga forskningsdata, som av erfarenheter från internationella projekt, är att man förmodligen bör närma sig systemfrågorna på ett pragmatiskt sätt och undvika att försätta sig i en situation där man låser sig vid ett, och endast ett systemval.

I vissa fall – främst vid hantering av enklare objekt där särskilda behov av att kunna kontrollera åtkomst och versioner inte föreligger – kan alternativ ett vara en fullt framkomlig väg. Samtidigt som alternativ två förmodligen på sikt utgör den mest ideala systemlösningen för merparten av de objekt som kan vara aktuella.

APPENDIX



Bilaga 1 - Forskningsprocessen 1

Forskningsdata inom konstnärliga och humanistiska vetenskaper – open access?

Intervjufrågor

Inledning

Projektets syfte

Kontaktvägar

Ev övergripande skiss.

Accepterar att deltar i projektet (kartläggning och leverera data).

Slutrapport.

Uppgifter om informanten

Namn, tjänstetitel, Institution, Universitet

Inledning

Forskningsprojekt

Vilket är ditt senast (alternativt största) publicerade forskningsprojekt där du använt forskningsdata?

- typ av data
- format och programvaror
- var det "typiska" data för ditt forskningsområde?

Fråga 1

Insamling och bearbetning av forskningsdata

- hur och varifrån hämtades primärdata och sekundärdata
- bearbetning
- analys
- registrering/lagring av forskningsdata under denna fas
- dokumentation (t ex procedurer för datainsamling, anvisningar, uppgifter om upphovsrätt, sekretess)
- problem

Fråga 2

Lagringsmedium

- hur lagras dina forskningsdata (t ex server) över tid

Fråga 3

Hur publiceras/publicerades av resultatet av forskningsprojektet

- tidskriftsartikel, rapport, bok, konferensbidrag, konstnärligt arbete m m
- med eller utan additional files – vilken typ av data? Forskningsdata?
- fritt tillgängligt - Open Access ^{30*}

Fråga 4

Tillgängliggörande av forskningsdata

- tillgängligt men med begränsningar (t ex enbart för andra forskare)
- hur viktigt är fritt tillgängliggörande
- har du haft kontakt med lärosätets öppna arkiv
- har du haft kontakt med Svensk Nationell Datatjänst (SND)
- har du haft kontakt med något ämnesarkiv

Fråga 5

Återanvändning av forskningsdata

- hur användbara är dina insamlade data för återanvändning för dig själv/andra forskare
- vilken typ av bearbetning skulle krävas
- hur lång tid tar avidentifiering, dokumentation m m

Fråga 6

Spridning av forskningsdata

- länkning mellan olika arkiv
- finns ämnesarkiv inom ditt forskningsområde
- länkning mellan publikation och tillhörande forskningsdata

Fråga 7

Hjälp

Saknar du någon typ av tjänst för hjälp med insamling, publicering, tillgängliggörande och spridning?

Fråga 8

Ditt deltagande i projektet.

Vill du låta oss hjälpa till att sprida och tillgängliggöra dina forskningsdata?
Eventuella förbehåll?

Tack!

[Bilaga 2, Intervjumall](#)

Hej [forskarens namn],
[lärosätets namn] deltar i ett projekt, där vi under våren vill komma i kontakt med dig som forskar inom humaniora och konstnärliga vetenskaper, i syfte att öka kännedom om och förhoppningsvis också tillgängligheten till dina och dina kollegors rå-/forskningsdata*.

Vi undrar därför om vi kunde få träffa dig under en timmes tid för att på ett strukturerat vis höra mer om ditt senaste (eller största) projekt där du haft ett rådataunderlag, och intervjua dig om din syn på var och hur dessa forskningsdata lagras och i vilken mån de kan göras tillgängliga för andra.

Vi kommer inledningsvis berätta om projektet**, för att därefter be dig berätta om ett av dig utvalt forskningsprojekt. Därefter utifrån projektets frågeställningar diskutera vidare med dig kring dess forskningsdata: dess typ (t.ex. format text/film); hur inhämtning sett ut (bearbetning/dokumentation); lagring (hårddisk/papper/annat); publicering (med/utan s.k. "additional files"); samt spridning och återanvändning (när/vem/hur kan dessa data valideras).

Du kommer träffa mig och [din kollegas namn]. Vi arbetar med den elektroniska publiceringen i [öppna arkivets namn], vi undrar också hur/om du skulle vilja se kopplingar mellan dina publikationer och dess forskningsdata där eller någon annanstans?

Har du möjlighet att träffas [datumförslag etc]

Med vänliga hälsningar,

[ditt namn]

* i projektet följer vi OECDs definition om research data "as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that commonly accepted in the scientific community as necessary to validate research findings" OECD, Principles and Guidelines to research Data from Public Funding, 2007

** Projektet "Forskningsdata inom humanistiska och konstnärliga vetenskaper - open access?" är en del av det nationella programmet openaccess.se och i projektet samarbetar vi med Svensk Nationell Datatjänst, Göteborgs universitet, Lunds universitet och Linköpings universitet [byt]. Projektet leds av Göteborgs universitet som också har gjort projektets hemsida om du vill titta lite mer:
<http://www.ub.gu.se/forskningsdata/>

Bilaga 3, Introducerande brev till utvalda forskare