

Slutrapport för projektet Långsiktig öppen tillgång till forskningsdata vid HT-fakulteterna och S-fakultet, Lunds Universitet

1. Projektets bakgrund och syfte

Under år 2016 har Riksbankens Jubiléumsfond samt HT-fakulteterna och S-fakultet vid Lunds Universitet samfinansierat ovan rubricerade pilotprojekt. Projektansvarig har varit professor Samuel Rubenson och projektledare forskningsingenjör Johan Åhlfeldt. I styrgruppen har dessutom ingått docent Matthias Baier, chefsbibliotekarierna Viktoria Hörnlund, HT-biblioteket, och Karin Jönsson, Samhällsvetenskapliga biblioteket. I projektet har även bibliotekarierna Annakim Eltén, Aron Lindhagen och Ann-Sofie Zettergren medverkat.

Projektet har handlat om hur forskningsdata (digitalt forskningsmaterial) ska bevaras och hållas tillgänglig på kort och lång sikt. Syftet är att förslå lösningar för hur vi inom de nämnda fakulteterna ska spara och hålla insamlat material, analyser, verktyg, webplattformar och databaser från forskningen tillgängliga efter att projekten är avslutade. Det övergripande målet har varit att ge förslag till hur Humanistiska och teologiska fakulteterna (HT) samt Samhällsvetenskapliga fakulteten (S) ska kunna skapa bättre förutsättningar för sina forskare att hantera, arkivera och långsiktigt tillgängliggöra sina forskningsdata. I korthet skulle projektet

- med utgångspunkt i pågående forskningsprojekt som använder digitala metoder undersöka vilken organisation, kompetens och vilka metoder som krävs för en forskningsnära löpande förvaltning och tillgängliggörande av forskningsdata inom HT- och S-fakulteterna.
- i samverkan med fakulteternas IT-enheter utreda vilka tekniska plattformar som bör ligga till grund för öppen tillgång till forskningsdata inom HT och S, liksom hur dessa kopplar till existerande och planerade strukturer på högre nivåer inom LU och nationellt.
- ge förslag till hur kompetensen inom fakulteterna kan förstärkas vad gäller digitala metoder inom forskning och undervisning.
- i samverkan med aktörer inom och utom LU (fakultets- och universitetsbibliotek, arkivansvariga, juridiska enheten, riksarkivet, Svensk Nationell Datatjänst (SND), SWECLARIN m.fl.) medverka till utformande av tydlig ansvarsfördelning och koordinering av insatser inom området.

Pilotprojektet är en fakultetsnära fortsättning på Lunds Universitetsbiblioteks tidigare, av RJ finansierade, utredning.¹ Det pågår även en rad andra initiativ inom forskningsdatahantering vid Lunds Universitet. På uppdrag av vicerektor har det även tillsatts en arbetsgrupp för forskningsdatafrågor med representanter från bl.a. Humanistlaboratoriet.

2. Förutsättningar och genomförande

Fakulteternas forskare och doktorander svarar idag själva för sin datahantering, arkivering och tillgängliggörande av forskningsdata. Ett relativt litet antal forskare och doktorander inom HT- och S-fakulteterna använder sig av SND:s tjänst för datapublicering. Flera forskningsprojekt vid HT-fakulteten använder via Humanistlaboratoriet en korpusserver för avancerad datahantering. En del forskare eller forskargrupper publicerar data med öppen tillgång på egna plattformar och ett nytt projekt avser att skapa en interaktiv publiceringsplattform i samarbete med Humanistlaboratoriet, Historiska museet och arkivet vid Lunds universitet. Gemensamt för dessa projekt är att de publicerar forskningsdata med den uttalade målsättningen att öppet tillgängliggöra denna för mycket lång tid framöver. Det är idag i praktiken i hög utsträckning forskarens och handledare/doktorandens eget ansvar att lagar och regler efterföljs och att data arkiveras.

Inom fakultetsbiblioteken arbetar flera personer specialiserade på forskningsstöd. T.ex. har biblioteken ett arbete med stöd för datahantering i form av seminarier kring olika aktuella teman för forskare och i samband med forskningsansökningar. Forskningsstöd erbjuds i olika grad form av information om det juridiska regelverket, datahanteringsplaner mm.

IT-enheternas nuvarande arbete med forskningsdata rör främst lagring, datasäkerhet och backup. Denna infrastruktur är knuten till den enskilde forskaren eller doktoranden under anställningen. Vid HT-fakulteten tillhandahåller IT-enheten också en tjänst för administration av virtuella servrar som ett antal stora forskningsprojekt vid HT-fakulteten utnyttjar. För S-fakultet fungerar Lunds Datacentral (LDC) som IT-enhet.

Humanistlaboratoriet är en tvärvetenskaplig forsknings- och träningsenhet vid HT-fakulteterna med användare från alla fakulteter vid LU och även från andra lärosäten. Humanistlaboratoriet ger tillgång till olika typer av utrustning och erbjuder forskningsförberedande kurser och introduktioner kring utrustning, metod och hantering av forskningsdata, t.ex. med hjälp av laboratoriets korpusserver. Humanistlaboratoriet är också en nod i forskningssamarbetet SWE-CLARIN² tillsammans med åtta andra universitet och organisationer. I samverkan med HumLab vid Umeå universitet pågår ett arbete med att etablera en gemensam metod- och kompetensförsörjning och säkerställa långsiktighet för interaktiva plattformar för publicering, visualisering och analys av forskningsdata online. Humanistlaboratoriet har f.n. en egen infrastruktur på serversidan, i samarbete med LUNARC.

Projektet har arbetat med en detaljerad processanalys för att identifiera behov av befintliga stödfunktioner för datahantering. Analysen har utgått från forskningsprocessen, från projektplanering och projektstart, till datainsamling, databearbetning, dataanalys, arkivering och publicering av data för återanvändning. För att få en bild av behoven hos forskare och doktorander har projektet intervjuat och samtalat med forskare och forskningsunderstödjande enheter både i och utanför Lunds universitet, t.ex. avdelningarna för dokumenthantering och juridik samt forskningsetik. Projektet anordnade i maj en workshop med fakulteternas forskare för att belysa problem och möjligheter inom datahantering samt inhämta erfarenheter och önskemål. Projektet har också haft möten med verksamhetsansvariga på LDC, HT-fakulteternas IT-enhet och Humanistlaboratoriet, deltagit i möten med forskargrupper kring avancerad datahantering. Med berörda parter har vi sett på i vilken omfattning LUCRIS-systemet kan användas för beskrivning av publicerade forskningsmaterial.

Projektet har även deltagit i det pilotprojekt som SND samtidigt bedrivit vid fem lärosäten i Sverige. SND:s syfte har varit att undersöka förutsättningarna för en kompetensöverföring i datahantering utifrån en modell för hur forskningsdata skulle kunna göras tillgängliga och långtidsbevaras vid lärosätena samt att undersöka SND:s framtida roll i detta som kompetenscentrum. I detta har personal från fakultetsbiblioteken vid HT-fakulteterna och S-fakultet, IT-enheten vid HT-fakulteterna, LDC, Humanistlaboratoriet, avdelningen för dokumenthantering vid LU och Malmö högskola, samt några av fakulteternas forskare deltagit. Ytterligare kompetensutveckling i olika former har genomförts inom ramen för projektet.

Projektdeltagare har därutöver medverkat i konferenserna *Mötesplats Open Access (MOA)* i Stockholm, konferensen *Digital Humaniora i Norden* i Oslo, SND:s diskussionsforum *Tillgång till forskningsdata – Finansiärernas krav och lärosätenas beredskap* i Göteborg, konferensen *Infrastruktur för forskningsinformation* i Sverige (IFFIS), *International Digital Humanities Symposium*, workshopen *Geokodat kulturarv och kartapplikationer*. Projektet har även gjort ett studiebesök vid Digital Social Science Lab i Köpenhamn och en projektdeltagare har gått kursen *Data Scientist Training for Librarians (DST4L)* vid Danmarks tekniske informationscenter i Köpenhamn. Projektledaren ingår tillsammans med forskare från HT-fakulteten i en referensgrupp för Alvin-portalerna för att bevaka möjligheterna för forskning på kulturarvsmaterial från universitetsbiblioteken i Sverige. Konsultationer har också skett med Digisam vid Riksarkivet som ansvarar för samordningen av digitaliseringen av bibliotek, museer och arkiv i Sverige.

3. Övergripande förslag

Behovet av stöd för datahantering för forskningen inom fakulteterna är svårt att överblicka. Potentialen för bruk av digitala metoder liksom ökad forskning med digitala material är stor. Behovet av arkivering av forskningsdata är akut eftersom det f.n. inte finns någon central arkivlösning. Medan

många forskningsprojekt kan tillgängliggöra datamaterial via SND är det brådskande att säkerställa långsiktig publicering av databaser, särskilt för de projekt som har aktiva system online.

Från forskarnas sida har uttryckts behov av mer kunskap om vilka lagar och regler som gäller för forskningsprocessens olika moment, från datainsamling till arkivering och publicering av data, inte minst vad gäller skydd av känsliga data från kvalitativa undersökningar, samt tydliga instruktioner från verksamhetsansvariga och fakulteterna vad gäller hantering av forskningsdata utan att detta utökar administrativa arbetet. Likaså efterfrågas bättre överblick över vilka digitala datamängder som finns i omvärlden tillgängliga för bearbetning och analys och kunskap om hur man effektivast använder dessa, samt bättre kunskap om och enklare tillgång till verktyg för insamling, bearbetning, analys, visualisering och publicering av data.

Inom fakulteterna finns god kompetens inom datahantering och publicering, men denna är inte samlad. Det gäller för fakulteterna att säkerställa att den kan utnyttjas inom forskning och undervisning. För detta krävs en forskningsnära rådgivning och utbildning vad gäller avancerad databehandling och bättre och tydligare rutiner för samverkan mellan enheterna för forskningsstöd och forskningsenheterna samt former för kollegial samverkan och internationellt utbyte.

Med utgångspunkt i projektets syfte och genomfört arbete av befintliga strukturer föreslår styrgruppen följande ansvarsfördelning och uppdrag:

1. Långsiktig arkivering och öppen tillgång till publicerat datamaterial och databaser från forskningen efter att projekten är avslutade

Arkivering

Arkiveringen av fakultetens forskningsdata är ett akut problem som för att leva upp till gällande lagstiftning måste lösas i samverkan med universitetsarkivet och universitetet centralt. I avvaktan på en central arkivlösning bör fakulteterna överväga att själva temporärt lagra de forskningsdata och andra digitala forskningshandlingar som skall bevaras enligt Lund universitets dokumenthanteringsplan. Detta förutsätter att det finns en beslutsprocess som omfattar verksamhetsansvarig, forskare/doktorand och arkivet och som avgör vad som skall bevaras och hur länge. Fakulteternas IT-stöd bör få i uppdrag att ta fram en temporär lösning i samarbete med fakultetsbiblioteken och avdelningen för dokumenthantering. Därmed säkerställs att forskningsdata bevaras för eftervärlden oavsett doktorandens eller forskarens fortsatta anställning vid universitetet. Arkivering av digitala forskningsdata är obligatorisk enligt gällande lag och tillhör den grupp av material som får gallras först efter tio år och som i många fall skall bevaras för alltid.

Hårdvara

Eftersom kostnaderna för den hårdvara som krävs för arkivering av digital publicering av forskningsdata gäller en fast infrastruktur avsedd att permanent bevara forskningsdata är det inte rimligt att dessa kostnader belastar fakulteterna. Vi menar att universitetet centralt verkar för att tillräcklig hårdvara för datahantering, arkivering och publicering ställs till forskarnas förfogande utan att detta belastar forskningsanslag eller verksamhet.

Rekommenderad programvara för datahantering

IT-enheterna bör tillsammans med fakultetsbiblioteken tillhandahålla en lista på rekommenderade programvaror för olika moment inom datahanteringsprocessen, ge support på dessa programvaror och föreslå verktyg som svarar mot forskarens behov. För att möta behoven bör IT-enheterna dessutom ha ett tydligt uppdrag att bistå forskarna med datahantering, t.ex. automatiserad datainsamling, samkörning av data från olika datakällor och transformation mellan olika dataformat. IT-enheten på HT-fakulteten erbjuder redan en tjänst som innebär förvaltning av virtuella servrar för långsiktig datapublicering. Utformningen av denna tjänst bör formaliseras så att ansvarsfördelningen mellan forskare och IT-enhet blir tydlig. Forskningsprojekten bör ansvara för utveckling och underhåll av applikationsmjukvara under projektiden medan IT-enheterna ansvarar för underhåll därefter. IT-enheternas ansvar bör omfatta uppdatering av operativsystem, backup och certifikat för säker anslutning. LDC erbjuder en liknande tjänst.

2. Dokumentation av, information om och grundläggande vägledning kring fakulteternas forskningsdata och datahantering

Publicering och registrering av forskningsdata

Enskilda forskare och doktorander intresserade av att publicera forskningsdata som *Öppna data* ska kunna påräkna stöd från fakultetsbiblioteken. En rekommendation är att publicerade data görs öppet tillgängliga med Creative Commons erkännande licens (CC BY) i anslutning till god forskningssed. En sådan licens kommer inte heller i konflikt med bestämmelser om allmänna handlingars offentlighet i Tryckfrihetsförordningen. Prefekt och forskare, handledare och doktorand, måste dock säkerställa att en sådan publicering av forskningsdata inte står i konflikt med bestämmelser i Personuppgiftslagen, Offentlighets- och sekretesslagen, Lagen om forskningsetisk prövning samt Upphovsrättslagen.

En ytterligare rekommendation är att publicerad forskningsdata beskrivs i modulen för datamängder i LUCRIS när den blir tillgänglig. Det kommer inte att bli möjligt att publicera forskningsdata inom LUCRIS, men däremot att länka till tjänster där data finns publicerade. En framtida registrering i LUCRIS av publicerad forskningsdata möjliggör en koppling till både forskare och publikationer, samt att datamängder kan redovisas på högre organisatorisk nivå, t.ex. institutions- och fakultetsnivå, och efter UKÄs forskningsämne. LUCRIS är dock inte ett system för administration i samband med datahantering, t.ex. tillämpning av sekretess, personuppgiftsbehandling, forskningsetisk prövning, datahanteringsplaner för pågående forskningsprojekt mm.

Lokal stödenhet för datahantering

Erfarenheter från flera forskningsprojekt visar på behovet av en tydligare samverkan och utökning av stödet för datahantering för att säkerställa bevarande och tillgänglighet till forskningsdata, liksom en tydligare ansvarsfördelning. Fakultetsbiblioteken bör lämpligen utgöra first-line support gentemot verksamhetsansvariga, forskare, handledare och forskarstuderande samt utgöra stödenheter för datahantering i nära samverkan med IT-enheterna och forskningseenheter såsom Humanistlaboratoriet. Ju tidigare en lokal samverkan kan komma in i datahanteringsprocessen desto effektivare kan resurser utnyttjas, dubbelarbete undvikas och kostnader minskas.

Webbguide

De inom projektet utarbetade webbguiderna för datahantering (en för HT-fakulteterna och en för S-fakultet) vänder sig till fakulteternas verksamhetsansvariga, forskare, handledare och doktorander med information anpassad till behoven på respektive fakultet. I dessa finns en egen ingång för målgruppen prefekter eller motsvarande som har ett särskilt ansvar för att regelverket följs, att data arkiveras och att data publiceras om forskare och doktorander blir bundna av sådana bestämmelser i framtiden. Vi föreslår att denna guide bör fortsätta att utvecklas och uppdateras av fakultetsbiblioteken. Avsikten är att all nödvändig information relaterad till datahantering ska vara synlig genom denna guide eller att det finns hänvisningar till informationsresurser hos andra enheter inom LU (dokumenthantering, juridik och etik) och externt, t.ex. hos SND.

3. Metodutbildning och metodutveckling inom datahantering i syfte att främja fakulteternas forskning

Utbildningsinsatser

De ökade kraven på datahantering, liksom de möjligheter som digitala metoder erbjuder, skapar ett generellt behov av ökad kunskap hos fakultetens forskare och doktorander. Eftersom doktorander är en nyckelgrupp för att höja den generella datahanteringskompetensen vid fakulteterna bör fakulteterna se över möjligheterna att i forskarutbildningen ge grundläggande utbildning inom datahantering och digitala metoder och på olika sätt tydliggöra de resurser inom datahantering som finns vid fakulteterna, vid LU centralt och nationellt. För dem som har intresse och behov av mer avancerad utbildning är det önskvärt att det finns möjlighet till fristående kurser i datahantering och i digitala metoder, gärna i samverkan mellan de berörda fakulteterna.

Avancerad datahantering och digitala metoder

Genom Humanistlaboratoriet har fakulteterna en god miljö för utveckling av avancerade metoder för insamling, bearbetning, analys och dynamisk publicering av forskningsdata. För att bättre utnyttja denna, få en större samverkan inom metodutveckling och ett ökat såväl nationellt som internationellt utbyte inom digitala humaniora bör externa medel som avser utveckling av digitala metoder inom projekt i ökad utsträckning läggas vid laboratoriet. Därmed skulle den samordning som redan sker inom språkvetenskapen genom SWE-CLARIN-samarbetet kunna utvidgas till andra områden och innefatta andra miljöer för digital forskning som t.ex. Digital Humaniora initiativet vid Växjö universitet³ och HumLab vid Umeå universitet⁴. Laboratoriet utnyttjas redan av forskare från Samhällsvetenskapliga fakulteten och med en ökad samverkan med forskare från andra delar av HT-fakulteterna kan laboratoriet få större betydelse, t.ex. genom analys och visualisering av geokodat material eller data som beskriver sociala nätverk. Den kursverksamhet vad gäller digitala metoder som idag sker vid Humanistlaboratoriet kan i framtiden behöva utökas. Genom ett utökat användande av Humanistlaboratoriet skulle en del dubbelarbete vad gäller metodutveckling och avancerad datapublicering kunna undvikas och det skapas bättre förutsättningar för att utvecklingsarbetet i olika forskningsprojekt berikar varandra och tas till vara för kommande forskning inom fakulteternas områden.

1 *Research Libraries and Research Data Management within the Humanities and Social Sciences* Johan Åhlfeldt & Johnsson, M. 2015. Lund University Library. Projektet var initierat av forskningsprogrammet Early Monasticism and Classical Paideia (RJ).

2 <https://sweclarin.se>

3 <https://lnu.se/forskning/sok-forskning/digital-humanities/>

4 <http://www.humlab.umu.se/sv/>