

1 **Intervjuare:** Jag tänkte börja med att berätta om vad det här projektet är och hur
2 intervjun går till. Tack så jättemycket för att du vill vara med! Syftet med det här projektet
3 är att kartlägga hur forskare arbetar med forskningsdata idag och vilka behov som finns.
4 Målet är att kunna utveckla det bästa möjliga stödet för forskare så att ni kan fokusera
5 på er forskning. Samtalet idag kommer att struktureras kring en förenklad modell av
6 arbetet med forskningsdata. Och jag ska visa hur den ser ut. Ser ni bilden nu?

7 **Antecknare:** Ja.

8 **Deltagare:** Jag ser.

9 **Intervjuare:** Ja. Vad bra. Ja. Så här kan du se att modellen den innehåller olika faser.
10 Planera, skapa, samla in, analysera, tillgängliggöra, bevara och rapportera. Vi kommer
11 specifikt att titta på någonting som kallas för förmågor. Det har ingenting med en
12 persons förmågor att göra utan det är ett begrepp som beskriver processer och
13 informationsresurser, kompetenser, regelverk, IT-stöd och så vidare. Som relaterar till
14 den e-infrastruktur som forskningen använder. Alltså helt enkelt allt det som behövs för
15 att kunna hantera forskningsdata på ett bra sätt. Genom alla faserna av en
16 forskningsaktivitet. Så vi kommer inte att titta på alla processer inom forskning idag, utan
17 vi vill särskilt undersöka vilken e-infrastruktur som behövs för att hantera forskningsdata
18 på ett bra sätt.

19 Och när vi pratar om e-infrastruktur så kan det vara de här sakerna. [visar bild] Det här
20 är ingen bild som du behöver memorera eller så. Utan det här är bara en bild med
21 exempel på vad e-infrastruktur kan vara för någonting. På den lägsta nivån så kan det
22 handla om de system och nätverk som behövs. För att du ska kunna spara och arbeta
23 med dina data. Men det kan också handla om de infrastrukturer och verktyg som behövs
24 för att du ska kunna samla in eller analysera data. Det kan vara plattformar som du
25 använder för att samarbeta kring data. Eller för att publicera data. Och allra längst upp
26 så finns det olika tjänster för råd och stöd. Som du kanske använder eller har tillgång till.
27 [datastödsfunktion på lärosätet] till exempel. Om det handlade om [uppdrag] som
28 [antecknare] beskrev att han jobbar i. Det skulle kunna höra hemma där.

29 Och allt det här är e-infrastruktur. Och det kanske är inget som man tänker på i det
30 dagliga arbetet. Men man tänker på e-infrastrukturen när den inte finns tillgänglig eller
31 när den inte fungerar. För den behövs för att du ska kunna göra ditt arbete som forskare.
32 Så ha gärna den här bilden i åtanke under intervjun. Eftersom den representerar den
33 miljön som du som forskare rör dig i och behöver förhålla dig till. Intervjun kommer gå till
34 så här. Att vi kommer utgå från en forskningsaktivitet som du kommer att tänka på.
35 Kanske ett aktuellt eller nyligen avslutat projekt. För det brukar vara enklare att prata om
36 någonting konkret. Men du får också gärna säga saker som är mer generella eller saker
37 du kommer att tänka på. Jag slutar dela här. Så ska [antecknare] få visa en bild istället.

38 **Antecknare:** Så där.

39 **Intervjuare:** Ja, och det här är samma karta som jag visade tidigare fast nu finns det lite
40 utrymme att skriva i den. För vi kommer att prata utifrån det här hjulet med de olika
41 faserna i ett forskningsprojekt. Och så kan du berätta om hur du arbetar med data i de
42 olika stadierna. Men du behöver inte hålla det strikt till dem. Du kan prata fritt. Och under
43 tiden så kommer [antecknare] att anteckna kortfattat på bilden med hjulet. Och sen efter
44 det så stämmer vi av och ser om du tycker att vi har fångat det du har sagt. Och du kan
45 rätta till om något har blivit fel eller om det är någonting som saknas som du vill lägga till.
46 Och sen i slutet så kommer vi be dig att ange vilken av figurerna uppe i vänstra hörnet
47 där som du skulle sätta på helheten för de olika faserna. Glad för bra, neutral för nja och
48 sur för missnöjd. Och det handlar inte om du är nöjd med din egen forskning utan liksom
49 hur nöjd du är med e-infrastrukturerna och den digitala miljön. Och hur den har funkat
50 för dig i de olika faserna.

51 Och det kan kännas kanske obekvämt att markera väldigt mycket med missnöjd. Men
52 var så ärlig som du kan. Och som jag sa tidigare så spelar vi in ljudet under intervjun.
53 Och det kommer transkriberas och därefter slängs inspelningen. Det som analyseras är
54 transkriptionen av intervjun och kartan som vi antecknar i. Har du några frågor?

55 **Deltagare:** Alltså det varierar ju väldigt mycket mellan olika, mellan olika studier. Om jag
56 har en studie som är på något sätt kopplat till [det uppdrag] som sker här så är ju den
57 infrastruktur jag använder helt annorlunda än om jag, till exempel, skulle göra någon
58 genetisk studie och vara beroende av NCBI eller någon sån genetisk resurs. Så det blir

ju väldigt, väldigt(--)) Det är ju svårt att (.) Jag menar ett vanligt experimentellt, en vanlig experimentell studie (.) en studie som är kopplad till [uppdrag] här. Eller en mer genetisk studie har ju helt olika behov av infrastruktur. Så jag vet inte, är det något speciellt som jag ska, om jag ska ta något exempel, vad är ni mest intresserade av då?

Intervjuare: Ja, alltså egentligen får du välja själv. Du får jättegärna berätta om en genetisk studie, till exempel. Eller om du hellre vill berätta om något kopplat till [data från uppdrag] så får du också göra det.

Deltagare: Det jag håller på mest(--)) Det forskningsprojekt som jag håller på mest med nu, det är väl det och sen en studie som en exjobbare håller på med. Sen så är det (.) Så här i [insjö] så ligger det gamla nät, alltså tidsserier av gamla nät, ifrån(--)) De påbörjades någon gång på 50-talet för ungefär 70 år sedan. Så, vi har (.) varit nere i biblioteket, så då, det är ju ingen e-infrastruktur. Men vi har varit nere i biblioteket och tagit reda på vad det var för några nät. Och sen så har vi kontaktat en finsk nättillverkare som har tillverkat de här gamla näten. Och nu håller vi på och lägger dem. Och tanken med det här är då att vi ska försöka återskapa den här tidsserien av provfiskedata. Så att vi ska kunna säga någonting om antropogen påverkan.

Men om man tar det då. Så i planeringsstadiet så är ju det(--)) Ni får berätta, ni får säga om jag eller ska jag bara babbla nu?

Intervjuare: Vi kan ta ett stadium i taget. Jag kan ställa lite frågor kring det som jag vill veta. Men berätta du det som är viktigt liksom

Deltagare: Nej, men planeringsstadiet, då jobbar jag ju dels med gamla, gamla data. Som inte är digitaliserade utan som finns i vårt bibliotek här. Men jag jobbar också emot de databaser som vi har [databas] och [databas]. Och försöker jämföra de här olika delarna. Och komma till någon form av slutsats med vad som egentligen gjordes där för 70 år sedan. Så det är lite detektivarbete med båda. Jag tycker det funkar ganska bra. Databasen är ju full av fel. Vilket inte är så konstigt eftersom det var 70 år sedan som det här utfördes. Och det är väl ingen som egentligen har kollat på det sedan dess. Så, precis som med alla andra databaser så är ju felsökandet beroende på användandet.

87 Vad mer kan man säga om dem? (..) De är ganska personberoende här på [enhet på
88 lärosäte]. Det är någonting som oroar mig lite grann. För vi har en väldigt bra
89 databasansvarig nu. Men börjar väl han, är väl lite över 60. Och han är ju biolog i grund
90 och botten. Jag vet inte hur, vad som kommer hända när han går i pension. Och det är
91 ju någonting sånt där som man kanske, som ligger i bakhuvudet lite. Ja. Några frågor på
92 den? Men ganska, ganska glad är jag väl här. Tror jag.

93 **Intervjuare:** Hur(--) När du planerade det här datainsamlingen, hur gick det till då? Har
94 du liksom, i praktiken? Du har använt data som fanns i biblioteket. Har du sen liksom
95 sammanställt den i ett Excel-ark, eller hur? Eller i har du vilken typ av databaser, eller?

96 **Deltagare:** Alltså, det är utdrag ur en accessdatabas.

97 **Intervjuare:** Ja. Okej.

98 **Deltagare:** Det startar ju med (.) Så det där är (.) Det här skulle jag säga är på
99 planeringsstadiet. Skulle jag tro. Men det startar ju med utdrag ur den här
100 accessdatabasen. Där man letar efter sammanhängande tidsserier. Och sen så när man
101 hittat en sammanhängande tidsserie så kollar man om redskap och så vidare är, liksom
102 metodik ger intrycket av att vara tillräckligt standardiserad för att gå att upprepa. Så jag
103 kollar helt enkelt på utdrag i en accessdatabas. Jag använder inte Excel speciellt mycket.
104 Jag brukar använda CSV-filer. Jag tycker de känns mer stabila. I och med att
105 databaserna är ganska stora så brukar jag(--) Jag använder ju mycket R, det här
106 statistikprogrammet. Det går väl att använda Python och sånt också. Men jag gillar att
107 använda R för att få kunna tabulera upp datan på annat sätt. Och söka efter de här
108 tidsserierna i det. Så jag använder generellt R väldigt mycket till nästan allting. Och jag
109 är inte jätteförtjust i Microsoft-verktygen. Jag använder inte Excel i annat än nödfall. Eller
110 när jag måste använder jag Excel, men helst inte.

111 **Intervjuare:** De här accessdatabaserna, var är de sparade någonstans? Har ni egna
112 servrar eller är de sparade på [lärosätes] servrar? Eller?

113 **Deltagare:** Det vet jag faktiskt inte. Jag skulle tro att de är sparade på [lärosätes] servrar.

114 Men jag vet inte. Det är [databas]. Så det är ju en utav de databaser som vi har någon
115 slags [uppdrag] för. Så gissningsvis måste de ligga på [lärosätets] databaser. Skulle jag
116 tro. Men jag kan ha fel där.

117 **Intervjuare:** Och sen när du sammanställer dina egna CSV-filer. Så gör du det på din
118 egen dator då?

119 **Deltagare:** Ja, precis. Vi har väl någon sån här uppkoppling mot servern. Så att jag
120 lägger dem ju inte (.) i dokument. Eller jag lägger dem i dokument. Eller någon av de
121 mapparna som uppdateras mot servern. Men de ligger i min lokala mapp i alla fall.

122 **Intervjuare:** Vet du, har ni sökt någon finansiering för den här datainsamlingen?

123 **Deltagare:** Ja det har vi, det är (.) Vad heter de där medlen? [finansiering] och (.) Ja, jag
124 har glömt bort vad de här medlen heter. Det är något av de här [lärosäte]-medlen för...

125 **Antecknare:** Utvecklingsprojekt?

126 **Deltagare:** Ja, utvecklingsprojekt. Det är det. Tack.

127 **Intervjuare:** Har ni då sökt pengar specifikt för, kopplat till liksom datahanteringen,
128 databehandlingen?

129 **Deltagare:** Det mesta går till tillverkning av nät här.

130 **Intervjuare:** Ja.

131 **Deltagare:** Det är ju ett ganska (.) I det här stadiet så är det ju inte (.) jättemycket
132 datahantering i det här. Utöver den datahantering jag redan gör. Jag sitter ju och jobbar
133 med de här med liknande uttag för [uppdrag]. Så, till exempel, [rapporter] och [rapport]
134 och sådana där uppgifter. Så jobbar jag med liknande uttag. Så i sig så är ju inte
135 dataarbetet för just den här studien jättestort.

136 **Intervjuare:** Nej.

137 **Deltagare:** Än.

138 **Intervjuare:** Är det ett projekt som du gör själv eller har du samarbetspartners som du
139 delar data med?

140 **Deltagare:** Jag delar i, eller data ska ju upp i våran databas som är tillgänglig. Men vi
141 arbetar med datat(--). Insamlingen av datat sker utav våran datainsamlingsenhet. Så det
142 är de som utför provfiskena. Och sen registrerar de även det här datat och lägger upp
143 det i våran databas.

144 **Intervjuare:** Så allt(--). Hur samarbetar ni kring data? Är det(--). De lägger in det i
145 databasen på eran server och sen så hämtar du data därifrån?

146 **Deltagare:** Ja, precis.

147 **Antecknare:** Jobbar ni, förlåt [intervjuare], jobbar ni mycket med fältapplikationer? Alltså

148 **Deltagare:** Vad sa du?

149 **Antecknare:** Jobbar ni också med(--). I insamlingsfasen där, alltså jobbar ni mycket med
150 fältapplikationer? Alltså pads eller (.) ?

151 **Deltagare:** De har något som de kallar för HD-fish. Det är någon form av sample size
152 mjukvara. Alltså att de delar in fiskarna i grupper för att mäta, längdmäta tillräckligt
153 många individer. Till exempel, om vi vill ha ett(--). Vi är intresserade av fördelningen av
154 storlekar här. Men i vissa fall så blir det väldigt många fiskar som fångas. Då har de en
155 slags fältapplikation som (.) som, där de matar in längdmätningar och som på något sätt
156 indikerar när de har fått tillräckligt många längdmätningar. I realiteten använder vi inte
157 den i de här fiskena eftersom de här näten enbart fångar större fiskar som vi inte har lika
158 stort stickprov på. Så, till den här studien(--). Jag menar de har med sig den ut. Men vad
159 jag vet så har de inte använt det än. Och jag tror inte de kommer använda den.

160 **Intervjuare:** Om vi går tillbaka bara lite kort till planeringsfasen. Upplevde du några
161 problem kopplat till data i den fasen?

162 **Deltagare:** Ja. Ja, alltså det är (.) Det var ju det som jag var inne på. Databaser är
163 beroende av att de används för att hitta fel i dem. Och det här datat är jättegammalt och
164 har inte använts sedan(--). Jag menar sedan det samlades in. På 60 (.) jag menar den
165 senaste användningarna av det är väl typ på 60-talet någon gång. Och därför så är det
166 väldigt mycket fel i det. Men det är ju--

167 **Intervjuare:** Hur hanterade du det? De här felen som du såg?

168 **Deltagare:** Det var ett provfiske som vi inte kunde göra. På grund av att det var(--)
169 Nättyperna var två olika nättyper var registrerade som samma nättyp. Men det är ju så
170 det är. Man får försöka felsöka och försöka pussla ihop det. Rent generellt så är det ju(--)
171 Ju fler som jobbar med en databas ju bättre fungerar den ju oftare och ju mer
172 korrigeringar blir inrapporterade.

173 **Intervjuare:** Ja. Okej. Vet du om du tog hjälp av någon stödfunktion i planeringsfasen?

174 **Deltagare:** Vad är en stödfunktion?

175 **Intervjuare:** Ja, hjälp med att skriva en datahanteringsplan eller hjälp från (.) Med någon
176 form av–

177 **Deltagare:** Alltså det faller väl under samma datahanteringsplan som vi har för alla våra
178 provfisken. I och med att vi har ett [uppdrag] så finns det ju en hanteringsplan för det
179 datat där (.) Efter (--) Det är ju då på skapandefasen, att efter att de har samlat in datat,
180 så går det till någon som kollar igenom att det ser korrekt ut. Och sen så rapporteras det
181 in till databasen och han som är ansvarig för databasen kollar också att data ser korrekt
182 ut. Och sen så rapporteras det in i databasen och blir tillgängligt.

183 **Intervjuare:** Ja, okej.

184 **Deltagare:** Så det finns en datahanteringsplan.

185 **Intervjuare:** Har du varit med och tagit fram den datahanteringsplanen eller det är nån
186 annan?

187 **Deltagare:** Nej, den har nog funnits här ett tag. Jag är ju själv inte riktigt involverad i
188 skapandeprocessen utan mer i planerande och analyserande processen.

189 **Intervjuare:** Ja, just det. Är det något mer som du kan berätta om själva
190 skapandeprocessen? Hur har ni tagit fram data? Du berättade lite grann om det.

191 **Deltagare:** Man har en utförandeplan. De sticker ut och lägger nät. Fångar fiskar.
192 Identifierar art. Väger och mäter dem. Och sen så... Jag är lite osäker på exakt hur de

193 registrerar datat sen. De kanske gör det i den här fältappen, det vet jag inte riktigt. Men
194 de skriver in vart någonstans i nätet fisken sitter. De identifierar art. Längdmäter och
195 väger den och registrerar det här datat då på individnivå. In i antingen en fältapp eller
196 om de använder Excel. Det vet jag inte riktigt.

197 **Intervjuare:** Har du upplevt några problem i den fasen kring datainsamlingen?

198 **Deltagare:** Nej, det är väl sådär. Ibland blir det fel. Det blir det alltid om man skriver in
199 väldigt många saker. Men det är inte så vanligt att man hittar större fel. Utan där brukar
200 de ofta(--) Han som har databasen har en del kontrollfunktioner som han kör på datat.
201 Så det(--) Jag tycker att det brukar vara rätt välkontrollerat. I alla fall de fiskena som sker
202 nu efter liksom att (..) Det har ju blivit betydligt mycket bättre med hur man rapporterar in
203 data. Tidigare har det varit väldigt mycket lösladsystem. Och då blir det ju mer fel om
204 man ska skriva in ett löslad.

205 **Intervjuare:** Ja, just det. Om du skulle berätta lite om hur du jobbar när du analyserar
206 data. När du har tagit ut de här data från databasen. Hur jobbar du med dem då?

207 **Deltagare:** Ja, jag har skrivit (.) Jag har skrivit en massa stödfunktioner som jag har på
208 en, våran restricted server, så en server som inte alla har tillgång till. Och det är
209 funktioner. Och sen så jobbar jag i R-skript där jag laddar ner de här R-funktionerna.
210 Och applicerar dem på datat. Så jag har en pipeline. En analytisk pipeline som jag
211 brukar köra. Som sker baserat på de här stödfunktionerna. Vi håller på och planerar och
212 migrerar det till GitHub istället. För att få det lite mer stabilt. Nu är det ju lite(--) Det här
213 funkar ganska bra men det är för att jag litar på alla i gruppen. Ifall att det börjar komma
214 in nån så här nyanställda exjobbare, kanske med mindre erfarenhet. Kanske de går in
215 och ändrar i funktionerna. Det skulle inte alls vara bra. Så, det är inte ett helt(--) Det är ju
216 inte, det blir mer stabilt om man har det till GitHub.

217 **Intervjuare:** Ja. I nuläget har du använder du R lokalt på din dator, eller är det?

218 **Deltagare:** Ja, bara lokalt.

219 **Intervjuare:** Japp. Då sparar du ner filerna på din dator och så jobbar du med
220 analyserna på din dator?

221 **Deltagare:** Ja, eller jag laddar filen ifrån en server. Så jag spar inte filerna på datorn
222 utan jag har(--) Analysen sker på datorn. Och sen så, ja, den sparas väl i RAM-minnet
223 antar jag. Men, jag spar inte ner den på datorn utan jag laddar den lokalt varje(--) Jag
224 laddar ner den.

225 **Intervjuare:** Finns det några fler typer av analyser som du gör?

226 **Deltagare:** Ja, det är ju ganska många olika analyser som utförs på datat. Men(--) Eller
227 vad tänkte du i skillnad ifrån?

228 **Intervjuare:** Ja, nej, du pratade nu om art och mätdata. Det är de analyserna som görs.
229 Det är inget mer?

230 **Deltagare:** Det görs ju ofta per art, antal, per art och nät. Och så kollar man då(--) Det
231 blir då ett relativt mått på abundansen. Så om du får 20 fiskar utav 20 abborrar ett år och
232 10 nästa år. Så är det möjligt att tänka sig att den har gått ner. Men sen så har vi ju även
233 andra variabler som ålder, längdfördelning, viktfordelning, kondition. Det vill säga vikt(--)
234 Skillnaden mellan den observerade vikten och en predikterad vikt för en given längd. Så
235 om, till exempel, det är dåligt med mat så kan fiskarna bli magra och då har de en längre
236 längd. Vikten är lägre än vad den borde vara för längden. Så det är ett ganska stort
237 antal variabler som används till det här.

238 **Intervjuare:** Samlar ni in någon annan typ av data? Till exempel, geografiska
239 koordinater för vad de är infångade, eller?

240 **Deltagare:** Inte just här.

241 **Intervjuare:** Vattenkemi eller någonting kopplat?

242 **Deltagare:** Inte just här. Eller vi har ju stödvariabler med temperatur och sånt. Men i och
243 med att de bara mäts vid provfiskena så är de inte så(--) De är inte så intressanta helt
244 enkelt. Jag tror att vi även tar syrekurvor och sånt där. Men det är lite svårt att använda
245 det i och med att vi inte har någon baseline för resten utav resten utav sjön. Men ibland
246 händer det att vi(--) Jag vet inte om vi kommer att göra det här, men eventuellt så
247 kommer jag att använda SMHIs databaser för att samla in väder. De har ju väldigt bra
248 väderdata. Det är ju inte på vattentemperaturer men det är på lufttemperaturer och

249 sådär, som man kan samla in och eventuellt jämföra lite mer om man vill korrigera för
250 om höstarna är varma eller kalla.

251 **Intervjuare:** Så i den här analysfasen av datahanteringen, har du upplevt några problem
252 där? Några svårigheter?

253 **Deltagare:** Alltså [lärosäte] är ju väldigt fästa vid att använda Windows. Jag är ju, jag
254 tycker ju att Unix mjukvara är jättemycket bättre för analys. Så jag har ju tidigare använt
255 Linux. Men nu verkar de inte vara så pigga på att man ska använda Linux. Jag tror att
256 det är någon säkerhetsgrej. Jag vet inte riktigt. Men det är ju(--) Windows är ju sämre på
257 i stort sett alla tänkbara sätt än Linux på analys. Och i och med att jag jobbar ganska
258 mycket med analys. Eller jag jobbar ju nästan bara med analys. Så är ju det lite
259 problematiskt. Det är också ibland när man(--) Det är någonting som inte stämmer med.
260 jag vet inte vad det är. Det är nåt som(--) Jag vet jättemånga som har problem med R
261 och [lärosäte]-datorerna. Att det liksom, det kan ta flera sekunder mellan att man trycker
262 på Enter och att analysen utförs.

263 **Intervjuare:** Ja jag har också haft... Jag känner igen det där faktiskt [skrattar]. Jag har
264 också haft lite problem med R. Det är någonting.

265 **Deltagare:** Det är jätte(--) Det är (--) Alltså när man sitter och analyserar flera tusen
266 rader kod. Då är det liksom ett problem. [skrattar]

267 **Intervjuare:** Jag förstår.

268 **Deltagare:** Och sen är det även de här säkerhetsuppdateringarna som känns som att
269 de nästan är ostoppbara. De kan ju(--) Ibland så kör jag analyser som tar väldigt lång tid.
270 Och då blir man ju inte glad om Windows får för sig att uppdatera sig utan att ställa
271 frågan i god tid. Så att man kan skjuta på det. Oftast kan man inte skjuta på det över en
272 hel analytisk period, liksom. Så det kan ju vara lite(--) Det är väl de saker som jag blivit
273 mest irriterad på. Det här med att det tar lång tid med R och de här
274 säkerhetsuppdateringarna. Det är väl de saker som har strulat till det mest för mig på
275 analysfasen. Då är ju mungiporna neråt. Då blir man ju upprörd.

276 **Intervjuare:** Vill du sätta den gubben redan nu? Annars kan vi ta det på slutet.

277 **Deltagare:** Annars är det väl ganska bra. Men på just de där sista grejerna så är det ju
278 en sur gubbe helt klart.

279 **Intervjuare:** Ja.

280 **Antecknare:** Får jag ställa en fråga där? Använder ni, eller använder du, de här mera
281 nationella plattformarna om, när det gäller UPPMAX till exempel, när ni har verkligen
282 stora eller hur säger man, alltså analytiska delar eller alltså CPU-tunga analyser som
283 ni ska utföra. Använder ni det också?

284 **Deltagare:** Jag har inte gjort det. Anledningen är att ofta så behöver man boka in tid och
285 sådär på de här analyserna. Och de är(--). För att man ska kunna använda de här
286 resurserna effektivt så bör analysen vara möjlig att parallellisera. Och mina analyser är
287 ofta inte jättelätta att parallellisera. Så det är möjligt att det skulle gå. Men de är inte
288 sådär uppenbara som till exempel(--). Många av de här genomiska analyserna är ganska
289 uppenbara att parallellisera. De här är lite mindre, de är mer liksom linjärt utförda. Jag
290 tror att om man var bättre på programmering än vad jag är så skulle det gå att
291 parallellisera. Men det är inte sådär uppenbart. Och kan man inte parallellisera dem då
292 är frågan vad det(--). Jag menar oftast är det ju inte CPU. Det är ju ofta inte processorn
293 som är begränsande. Utan det tar ju typ lika lång tid om man inte parallelliserar det.

294 **Antecknare:** Ja.

295 **Intervjuare:** Om man går över till tillgängliggörandet då. Har ni tillgängliggjort de här
296 data på något sätt för andra genom att publicera dem eller dela dem på något sätt?

297 **Deltagare:** Ja det ingår väl i vår uppgift att [ha ett uppdrag] för de här databaserna.
298 Så det sker ju (--). Det sker ju automatiskt. Det är ingenting som jag är involverad i utan
299 när det dyker upp på [databas] så är det offentligt.

300 **Intervjuare:** Har ni också publicerat någon artikel eller rapport?

301 **Deltagare:** Det här är helt nytt! Så här har vi inte publicerat någonting. Däremot så
302 kommer (.). Vi kommer att skriva rapporter om det i alla fall. Jag tror jag vet inte riktigt.
303 Jag har inte riktigt bestämt mig hur man ska göra med publicerandet på det. Jag vill
304 helst inte publicera innan det blir lite bättre. För problemet med (.) problemet med det

305 här är att man egentligen behöver ungefär tre år för att säga någonting. Och det sker i
306 [insjö] där det finns folk som är ganska missnöjda och tror att antalet fiskar har minskat
307 drastiskt och sådär. Så om man då publicerar den efter ett år, om det här är ett konstigt
308 år som avviker, så kommer folk att reagera väldigt(--). De kommer inte att läsa det här
309 finstiltat med att man behöver många år, bla, bla, bla, bla, bla, bla utan de kommer bara
310 reagera på resultatet. Jag vet inte, vi får(--). Man får ta och fundera lite på det där sen.
311 Men på något sätt så kommer vi att rapportera det. Och förhoppningsvis i vetenskapliga
312 artiklar. Men det kommer att dröja ett par år.

313 **Intervjuare:** Ja, okej. (.) Finns det några kostnader som kopplar till det här med
314 tillgängliggörandet i databaserna? Eller det ingår i den finansiering som ni har, antar jag,
315 eller?

316 **Deltagare:** Ja, som jag har förstått det så ingår väl de här(--). Det här ingår väl i
317 basfinansieringen till [lärosäte]? Att vi [har ett uppdrag] för det där. Men, jag kan
318 ingenting om det. Jag vet bara att det händer. Och att det finns kostnader som täcks
319 utav någon form av finansiering till någon.

320 **Intervjuare:** Har du upplevt några problem kopplat till att tillgängliggöra data?

321 **Deltagare:** Nej.

322 **Intervjuare:** Nej.

323 **Deltagare:** Jag tycker att det har blivit bättre. Vi har ju de här rapportserierna på [enhet
324 på lärosäte] nu som. Tidigare så fastnade ju väldigt mycket i grå litteratur, vilket ju är lite
325 olyckligt! Och det är ju redan(--). Det är även medan jag har jobbat här. Jag har bara
326 jobbat här i fyra år. Men när jag började så känns det som att man skrev väldigt mycket
327 grå litteratur som jag misstänker att ingen egentligen läste. Och väldigt få kunde läsa.
328 Jag tror väldigt få kunde läsa det. Och ingen utav de som kunde läsa det faktiskt läste
329 det. Och nu sker det ju(--). Nu rapporteras ju mycket av det här i sådana här
330 [publikationer] och [publikationer], vilket ju är intressant och möjligt att tillgängliggöra.

331 **Intervjuare:** Så hur tänker ni kring eller jobbar kring att bevara data långsiktigt?

332 **Deltagare:** Alltså datan bevaras väl i våra databaser? Så de finns väl dels i
333 papperskopia i vårt bibliotek eller i något arkiv här. Men även på då, digitalt. Analyserna
334 bevaras genom att R-kod tillgängliggörs. Och det har ju blivit väldigt mycket bättre nu
335 med artificiell intelligens och sådär man kan använda för att annotera den R-kod som
336 man har skrivit.

337 **Intervjuare:** Och hur tillgängliggör ni den då? Är det på GitHub eller?

338 **Deltagare:** Alltså, gissningsvis (.). Nu har jag inte publicerat någonting på det här. Men
339 fler och fler tidskrifter vill ju att analyserna ska finnas tillgängliga i någon form. Så då
340 som en *supplementary* materialfil. Och då blir det ju helt reproducerbart, resultaten. Man
341 kan ladda ner datat och man kan filtrera det och analysera det i R. Så det är ju en, ett
342 väldigt bra sätt, att bevara. Det är ju inte(--) Det är viktigt att det inte bara är datat som
343 ska bevaras utan det är också hur du analyserar det datat. Annars så(--) Båda de
344 stegen måste finnas med om man ska kunna tillgängliggöra hur olika slutsatser har
345 dragits.

346 **Intervjuare:** Ja. Nej men, om du skulle titta på det här som vi har skrivit nu. Tycker du
347 att det ser rätt ut? Eller skulle du vilja lägga till någonting?

348 **Deltagare:** Nej, jag tycker det ser bra ut!

349 **Intervjuare:** Så om vi skulle ta de här små gubbarna uppe i hörnet då. Planera-fasen.
350 Tyckte att du sa att du var glad gubbe på den eller?

351 **Deltagare:** Ja, den är väl ganska glad, tror jag.

352 **Intervjuare:** Skapandet?

353 **Deltagare:** Ja, den är väl också ganska glad, tror jag.

354 **Intervjuare:** Vad är det som du tyckte gick bra där? Eller vad är det, det funkar bra helt
355 enkelt eller?

356 **Deltagare:** Det är ju skönt att det görs ganska automatiskt. Jag vet inte. Jag tycker det
357 funkar och att det funkar är väl bra. Då blir man bara glad.

358 **Intervjuare:** Ja! Analysfasen då?

359 **Deltagare:** Alltså det är ju så stor del. Det här är ju typ (.) Säkert 80% av mitt jobb är i
360 analysfasen. Så det är ju svårt att ge en glad gubbe eller en ledsen gubbe. Jag tycker
361 väl att de här, sättet vi jobbar med att analysera saker tycker jag är ganska bra. Det
362 tycker jag, trivs jag rätt bra med. Så där kanske glad gubbe. Och sen så är jag inte så
363 glad över de här uppdateringarna och att man är ganska låst i vilka mjukvaror man ska
364 använda. (.) Men jag tycker vi får bra stöd här på i alla fall på [enhet på lärosäte]. Jag
365 tycker våran dataperson är duktig. Men det är ju inte någon analytiker och kan inte
366 använda de här programmen, vet inte vad det är och så.

367 **Intervjuare:** Har du tagit någon hjälp i med analysfasen eller IT-tjänster kring det?

368 **Deltagare:** Nej. Eller jag försökte fråga lite varför det här händer med R. Men ingen
369 verkar veta det.

370 **Intervjuare:** Ja. Nej, men så tillgängliggöra-fasen då, hur nöjd är du med den?

371 **Deltagare:** Ja, jag tycker det funkar bra. Jag är väldigt lite involverad i den. Men det
372 händer och det händer ganska fort och utan problem. Så det är väl glad gubbe.

373 **Intervjuare:** Hur funkar det med bevarandet av datan?

374 **Deltagare:** Det är väldigt likt tillgängliggörandet. Jag är inte involverad i det men det
375 händer automatiskt. För de här delarna får ni nog fråga någon av de som jobbar med
376 tillgängliggörandet och bevarandet.

377 **Intervjuare:** Vilka är det som jobbar med det?

378 **Deltagare:** [Namn] till exempel.

379 **Intervjuare:** Är det... Är det kollegor?

380 **Deltagare:** Ja, det är den som är ansvarig för [databas] här på... på [enhet vid lärosäte].
381 [talar långsamt]

382 (..)

383 **Intervjuare:** Och rapporteringen då? Hur känner du kring den?

384 **Deltagare:** Ja, den är ju ganska ny, de här rapportserierna. Jag tycker att det har blivit
385 bättre (.) Jag tycker det har blivit mycket bättre. Det är väl viss risk att vi konkurrerar lite
386 med(--) Det är även en granskningsfas i det där. Jag tycker att det har blivit mycket
387 bättre, men det har blivit bättre till den grad att jag undrar om vi konkurrerar lite med oss
388 själva. Det vore ju bättre att publicera det här som vetenskapliga artiklar. När det håller
389 för vetenskapliga artiklar. Nu är det ganska många utav de här rapporterna som kommer
390 ut ifrån våran avdelning. Som borde, som det nog hade varit bättre för alla om vi hade
391 publicerat det som vetenskapliga artiklar, istället för som rapporter.

392 **Intervjuare:** Men ni planerar också att skriva artiklar också så småningom, eller?

393 **Deltagare:** Ja, alltså jag vill ju skriva artiklar om det här. Och jag vill helst inte rapportera
394 det så mycket innan jag har skrivit artiklar. Men jag får se. Jag måste läsa vad jag har
395 skrivit i, vad jag har lovat. Det minns jag inte riktigt. Men har jag skrivit att det ska bli en
396 rapport så lär jag skriva en rapport. Men helst, med just det här projektet, så väntar jag
397 helst tills det tills jag har starkare data. Och det kommer inte vara i år.

398 **Antecknare:** Får jag flika in där lite? Är det inte så att syftet med rapporterna är lite mer,
399 till exempel, också att rapportera tillbaka till uppdragsgivaren, alltså [uppdragsgivare]
400 eller [uppdragsgivare]? Och sen, om jag har förstått det rätt, så är det ju oftast, det är ju
401 lite mer(--) Alltså data, själva data, publiceras inte genom rapporten. Det är ju mer en
402 sammanfattning eller en (.) Ja, eller en analys på något sätt. Det är inte så att data
403 själva publiceras eller, i samband med rapporten?

404 **Deltagare:** Nej, nej, det tillgängliggörs tidigare.

405 **Antecknare:** Ja, via databasen?

406 **Deltagare:** Ja. Alltså, det tillgängliggörs ungefär(--) Alltså, när jag får tillgång till det så
407 tror jag i stort sett alla har tillgång till det.

408 **Antecknare:** Men då i rapporterna lägger du inte heller in till exempel R-kodskripterna
409 eller så, det är mer som i en aggregerad form?

410 **Deltagare:** Ja, precis. Ja, utan, det är mer lite mer i artikelform. Men jag skulle inte
411 lägga till någon data eller något sånt där. Jag skulle nog inte lägga till någon R-kod

412 heller, just nu. Men det kan man ju göra om man vill det. Det är väl lite hur mycket tid
413 man lägger ner på det. Det blir ju väldigt många rapporter man skriver. Och man kan (.)
414 man måste spara lite tid på en del av de här kortare rapporteringarna som är mer (.)
415 som inte är slutgiltiga. Utan som inte är slutrapporter liksom utan som är någon form av
416 intermediär avrapportering. De kan man inte spendera så mycket tid på. För då har man
417 inte tillräckligt med tid att lägga på dem som ska slutrapporteras.

418 **Intervjuare:** Så hur(--). Om du skulle välja hur du känner för den fasen, rapporteringen?

419 **Deltagare:** Det är väl ganska glad gubbe, tror jag!

420 **Intervjuare:** Så slutligen, undrar jag om du skulle beskriva ditt drömscenario. Om du har
421 några önskemål för eller tankar kring hur det skulle kunna fungera om det fungerade
422 perfekt. Är det någonting som skulle kunna vara bättre? Och hur skulle det se ut? Kring
423 data och e-infrastruktur?

424 **Deltagare:** Alltså den stora grejen som stör mig är väl de här [lärosäte]-marinerade
425 Windows-datorerna. Det känns som att någon som jobbar med analys skulle behöva
426 engagera sig lite! Eller någon som kan mycket om analys och styra upp det där på något
427 sätt. Så att det går att (.) Så att det finns en medvetenhet om att konsekvenserna kan bli
428 ganska stora ifall att man lägger in en automatuppdatering, mitt i natten. Det är väl de
429 grejerna som jag känner, att de inte funkar så bra. Annars så tycker jag väl att det är
430 ganska, det funkar väl rätt bra, här.

431 **Intervjuare:** Ja, nej, men det var de frågor jag hade. Jag vet inte, har du något du
432 undrar, [antecknare] som du vill lägga till? Jag tror din mick är av. Jag vet inte om du sa
433 något.

434 **Antecknare:** Nej, det var allt. Inga ytterligare frågor.

435 **Intervjuare:** Hade du någonting som du ville lägga till, [deltagare]?

436 **Deltagare:** Nej, det är det är lugnt.

437 **Intervjuare:** Du får jättegärna höra av dig om du kommer på någonting i efterhand. Som
438 du skulle vilja lägga till.

439 **Deltagare:** Om jag kommer på något så säger jag till. Men jag tror inte det. Det känns
440 som att jag har babblat på ganska mycket. Det borde ha täckt det som...

441 **Intervjuare:** Ja. Om det är så att du har frågor som gäller datahantering och dataanalys
442 och dataplanering så finns ju vi i [datasstödfunktion vid lärosäte] som en stödfunktion.
443 Så du får jättegärna kontakta oss. Och om du har frågor som gäller [uppdrag] så kan du
444 kontakta [antecknare]. Så vi finns ju här och kan hjälpa till.

445 **Deltagare:** Ja, just det. Det känns ju som att det fungerar ganska bra. I och med att jag
446 inte har känt några sådana behov så är det ju, känns det som att det fungerar rätt bra,
447 som här.

448 **Intervjuare:** Ja. Men vad bra! Ja, det var allt jag hade. Så stort tack för att du ville vara
449 med.

450 **Deltagare:** Ja, tack själv och ha det bra.

451 **Antecknare:** Ja, tack så mycket [deltagare].

452 **Deltagare:** Kul att ses.

453 **Antecknare:** Ha det så fint.

454 **Deltagare:** Ha det bra.

455 **Intervjuare:** Hej.