

# Total röra runt hur forskare lagrar data

TEXT: SIV ENGELMARK

Mängden forskningsdata växer lavinartat. Men svenska universitet saknar i dag struktur för att säkra data för framtida analyser.

– Vi satsar miljoner på att generera data men saknar system att långtidslagra dem, säger Per Kraulis vid Science for Life Laboratory.



Leif Johansson.

I fjol kunde två forskare från Uppsala universitet presentera uppseendeväckande resultat om mikroplaster i vatten. De hade i försök sett att abborryngel åt de pyttesmå mikroplasterna hellre än sin naturliga föda vilket gjorde att en stor del av fiskarna dog. Resultaten publicerades i den ansedda tidskriften Science och fick stor internationell uppmärksamhet. Men kollegor som varit med på samma forskningsstation kände inte igen experimenten som de beskrevs. De anmälde forskarna för fusk. Och i april kom Centrala etikprövningsnämndens expertgrupp fram till att forskarna hade fuskat. Slutsatsen grundades bland annat på att de inte kunde redovisa bakgrundsdata. Uppsalaforskarna

förklarade det med att data fanns på en bärbar dator som hade blivit stulen.

– Det är en skandal att det kan vara en möjlig förklaring att data kan försvinna, säger Per Kraulis vid Data Office vid Science for Life Laboratory i Stockholm.

**Men förklaringen är** inte helt osannolik. I dag finns inte gemensamma rutiner för hur data som forskare tagit fram ska lagras. Universiteten har ett ansvar för arkivering, men det räcker inte. Det finns heller inget krav på hur länge data ska sparas. Därför kan den faktiskt komma bort.

– Vi satsar miljoner på att generera data men saknar system att långtidslagra dem, säger Per Kraulis.

På hans egen arbetsplats Science for

Life Laboratory gör man exempelvis storskaliga genstudier, som sekvenseringar av hela mänskliga genom, på uppdrag av forskare i Sverige. När de är gjorda skickas sekvenseringsdata till datorcentret Uppmax i Uppsala för analys. Där lagras den medan projektet pågår. Sedan tar forskaren med data till universitetet.

– Det är ytterst oklart vad som händer därefter. Någon lagrar på institutionens databank, eller egna hårddiskar. Doktorander och postdoktorer flyttar och ingen vet vart deras dataset tar vägen. Det är en total röra. Det finns data från projekt gjorda här för bara fyra, fem år sedan som vi har svårt att hitta.

Det här ställer till problem om någon exempelvis vill göra om en analys,



Den nya sekvenseringsmaskinen vid Science for life laboratory testas fortfarande men det går ändå nästan dubbelt så snabbt att sekvensera gener som med dagens maskiner. Det innebär att mängderna data ökar snabbt. Problemet är var den sedan ska lagras.

eller använda hopsamlade datamängder för helt nya analyser.

– Finns ett explicit beslut att slänga data så är det OK. Men det ska inte bara hända att de försvinner, säger Per Kraulis.

**Men nu kan en lösning** vara på väg. Sunet (Swedish university computer network) som sköter datanäten för såväl universitet som forskningsinfrastrukturer har börjat ta fram ett förslag.

– Vi diskuterar deras behov. Detta kom upp. Vi tittar på hur en lösning kan se ut tekniskt och försöker nu dra igång verksamhet inom området, säger Leif Johansson vid Sunet.

Behoven kommer att öka de närmaste åren när alla nya forskningsan-

läggningar är i full drift, som Science for Life Laboratory, synkrotronljusanläggningen Max IV, radaranläggningarna inom EISCAT och rymdobservatoriet Onsala.

– De fyra kommer att generera vassinniga mängder data. En grov kalkyl är att behovet av lagring kommer att öka med mellan fem och tio petabyte per år, mot dagens 0,5–1 petabyte per år. Jag tror det är en försiktig beräkning, säger Leif Johansson.

**Sunet skissar nu på** en nationell anläggning, gemensam för alla som producerar forskningsdata. Regeringen i Finland satsade för något år sedan på att bygga en sådan. Nu lagras forskares data på servrar som står i ett nerlagt pappersbruk i norra Finland.

– I Sverige kan tjänsten vara var som helst. Vi har sett till att det finns nät med hög kapacitet i hela landet. Vi har existerande avtal med universiteten som vi skulle kunna bygga vidare på. Men för många är frågan ny och inte färdigdiskuterad.

Enligt Leif Johansson blir det betydligt billigare för ett lärosäte med en gemensam lagring än om varje lärosäte hittar sin egen lösning. Det finns också stora energibesparingar att göra.

– Vi har ett problem som vi måste lösa gemensamt. Inget enskilt universitet klarar det men det går att lösa om alla forskningsinfrastrukturer plus universitet är med, till en kostnad lägre än hos molntjänstleverantörer, säger han. ■