

## Data för framtiden

### Pass 1: Inledning och bakgrund

*BAS Online 2021-01-20*

I den här presentationen får du en introduktion till den nationella infrastrukturen för bevarande och tillgängliggörande av svenska forskningsdata som håller på att växa fram. Vetenskapsrådet fick 2017 i uppdrag av regeringen att samordna det nationella arbetet med att införa öppen tillgång till forskningsdata. Målet är att en omställning ska vara genomförd senast 2026. Inom EU finns sedan tidigare en överenskommelse om att göra forskningsdata mer öppet. Detta innebär att lärosätena behöver ta ett större ansvar för sina data.

Svenska lärosäten har sedan tidigare skyldigheter för sina forskningsdata i egenskap av huvudmän och arkivmyndigheter, men hittills har arkivering av forskningsdata inte haft någon högre prioritet. Det här ser ut att förändras framöver, inte minst på grund av EU:s dataskyddsförordning som är svensk lag från den 25 maj 2018, men även på en förändrad inställning hos lagstiftare, forskningsfinansiärer och de förlag som publicerar forskningsresultat. Men att bevara data på ett sådant sätt så att de kan användas av andra i framtiden innebär att ett antal andra system måste finnas på plats.

Ett sådant system är utbildningar av forskare i datahantering och datahanteringsplaner. Man kan förvänta sig att datahanteringsplaner kommer att krävas i allt större omfattning av finansiärer, både i ansökningar om forskningsmedel och under själva projekten. Det behöver också finnas kontrollfunktioner på lärosätena för att säkerställa att de dataformat som forskningsdata bevaras i är så framtidssäkrade som möjligt, så att man inte riskerar att förlora data på grund av att filerna inte längre går att läsa. Kontrollfunktionerna behöver även se till att den dokumentation som medföljer data är tillräcklig för att någon annan kan förstå dem – en

granskare på en tidskrift, en forskare som vill replikera studien eller en forskare som vill använda data för en ny studie.

Det behöver också finnas processer och system som gör det möjligt att tillgängliggöra data, vilket inkluderar funktioner och system för att hitta och lämna ut data. Det räcker inte att bara bevara data; om de ska kunna utnyttjas maximalt så behöver de kunna sökas fram och laddas ned eller beställas på ett effektivt sätt, och då måste de processer och arbetsflöden som finns kunna garantera att de data som lämnas in av en forskare är desamma som lämnas ut till nästa forskare.

Det kan tyckas vara lärosätenas ansvar att dessa system kommer på plats, men i slutändan så måste systemen fungera inte bara på en lokal nivå utan även nationellt. Med andra ord måste de lokala insatserna fogas samman till en nationell lösning. Det behövs en svensk infrastruktur för att forskningsdata ska kunna göras lätt sökbara, tillgängliga och återanvändningsbara för alla svenska forskare och för att inte varje lärosäte ska tvingas hitta på egna lösningar på hur forskningsdata ska hanteras, bevaras och tillgängliggöras.

I SND:s ansökan till Vetenskapsrådets infrastrukturutlysning 2017 presenterade vi ett förslag på hur en sådan nationell infrastruktur skulle kunna se ut. I enkla drag går förslaget ut på att forskare får stöd från sin lokala DAU med att förbereda data för bevarande, tillgängliggörande och återanvändning. DAU:en granskar data och metadata, ser till att data lagras och skickar metadata till SND:s nationella metadatakatalog. De här metadata inkluderar information om var det finns en version som passar för förmedling av data till forskare och hur den är tillgänglig, dvs. om den kan laddas ned direkt eller om den måste beställas. SND stöttar också DAU:erna med kompetens och utbildning, och är drivande i det nationella DAU-nätverket, inte minst medan lärosätenas DAU-funktioner byggs upp. SND är också aktivt i samverkan med nationella och internationella intressenter och forskningsdatainfrastrukturer.

Lagringen av forskningsdata kan ske på olika sätt. Förutom disciplin-specifika nationella och internationella repositorer kan man tänka sig lokala möjligheter att lagra data, eller så kan lärosäten hyra egna delar av en stor nationell lagringsyta. Man kan även tänka sig att riktigt stora dataset lagras där de skapas för att det blir för krävande att flytta dem till andra lagringslösningar. Oavsett vilken lagringslösning eller kombination av lösningar som ett lärosäte använder sig av så behöver metadata för dataseten innehålla permanenta identifierare som talar om var data finns, så att en forskare som vill komma åt ett visst dataset kan göra det.

Jag ska inte fokusera så mycket på SND här – om du vill veta mer om hur SND fungerar så finns det mycket mer information på den här webbplatsen. Men det är värt att nämna att Svensk nationell datatjänst är ett nationellt konsortium som består av nio universitet: Chalmers tekniska högskola, Göteborgs universitet, Karolinska institutet, Kungliga Tekniska Högskolan, Lunds universitet, Stockholms universitet, Sveriges lantbruksuniversitet, Umeå universitet och Uppsala universitet. De här universiteten deltar alla i SND-konsortiet med olika specialiteter inom arbetet med forskningsdata. SND:s uppdrag är att arbeta för att svenska forskningsdata bevaras och tillgängliggörs för framtida forskning, och en viktig del av det här är att träna, utbilda och stötta DAU:erna och DAU-nätverket. SND jobbar därför med att hålla sig à jour med vad som händer på forskningsdataområdet, till exempel genom samarbeten med andra datarepositorier runt om i Europa, genom deltagande i EU-projekt som rör forskningsinfrastruktur och genom en aktiv omvärldsbevakning. Det vi lär oss försöker vi sedan sprida i DAU-nätverket. Vi arbetar också med att förbättra möjligheterna för att återanvända forskningsdata genom att bedriva ett aktivt utvecklingsarbete inom olika infrastrukturer, i samarbeten inom konsortiet och tillsammans med DAU-nätverket och olika parter inom det.

Det är viktigt att komma ihåg att medan data och metadata flyter i en cirkel från forskare till DAU:er till SND och tillbaka till forskarsamfundet så rör

sig kunskap åt båda hållen mellan de här tre aktörerna: varje grupp har något som de kan lära de andra, och något de kan lära av de andra. Till syvende och sist är den svenska forskningsdatainfrastrukturen avsedd att vara en hjälp för forskarna och att göra svensk forskning effektivare och bättre.

## Sammanfattning

Jag har i den här presentationen kort beskrivit hur DAU:ernas lokala roll och SND:s nationella roll förhåller sig i en svensk infrastruktur för forskningsdata. I den här infrastrukturen ingår lokala DAU:er som ger stöd till forskarna och vidarebefordrar data och metadata till lagring och tillgängliggörande via olika lagringsytor, samt en nationell metadata katalog som SND ansvarar för. I nästa presentation tar vi upp DAU:ens roll och arbetsuppgifter i större detalj.