

Utbildningsdepartementet

Riksarkivets synpunkter på regeringens forsknings- och innovationspolitik

Riksarkivet tackar för möjligheten att inkomma med synpunkter inför kommande forsknings- och innovationspolitiska proposition.

Sammanfattning

För att Sverige ska kunna möta samhällsutmaningarna krävs högkvalitativa forskningsinfrastrukturer. Riksarkivet förfogar över mycket stora mängder autentisk information och ligger också i framkant vad gäller utveckling av AI-modeller som stödjer datadriven forskning. För att ta tillvara på den stora potential som detta sammantaget medför föreslås följande åtgärder:

- 1) Ett tillägg i förordning med instruktion för Riksarkivet om att utgöra en forskningsinfrastruktur.
- 2) Höjt anslag om 15 Mkr för att utveckla Nationell Arkivdatabas som e-infrastruktur för forskning.
- 3) Höjt anslag om 30 Mkr för storskalig digitisering av samlingar i Riksarkivet inklusive infrastruktur och beräkningskraft för utveckling av AI-modeller.

Med dessa förstärkningar skulle Riksarkivet på ett konkret sätt bidra till att uppfylla regeringens forskningspolitiska mål om att Sverige ska vara ett av världens främsta forsknings- och innovationsländer och en ledande kunskapsnation.

Inledande synpunkter

Riksarkivets uppdrag är att säkerställa samhällets arkivinformation och göra den möjlig att använda över tid. Enligt arkivlagen ska myndigheternas arkiv bevaras, hållas ordnade och vårdas så att de bl.a. tillgodoser forskningens behov.¹ Riksarkivet ska verka för ökad kunskap grundad på forskning och i samverkan med andra förmedla kunskap inom sitt verksamhetsområde. I uppdraget ingår att tillgängliggöra arkiven för forskning och annat nyttjande, bl.a. genom digitalisering och tjänster för elektronisk tillgång.² Riksarkivet utgör i den meningen en forskningsinfrastruktur, något som

¹ § 3 Arkivlagen (1990:782)

² §§ 5 och 8 Riksarkivets instruktion (2009:1593)



också slås fast i *Stärkt fokus på framtidens forskningsinfrastruktur* (SOU 2021:65).³

Sverige har en lång arkivtradition, vilket har genererat stora mängder autentisk information som nu kan nyttjas inom många discipliner för att skapa samhällsnytta. De senaste årens utveckling inom digitala metoder och särskilt artificiell intelligens (AI) medför också nya möjligheter för forskning och arkivanvändning. Stora investeringar inom AI görs i vår omvärld och det är viktigt för Sverige att inte hamna på efterkälken. En utmaning för Riksarkivet, som forskningsinfrastruktur, är att de analoga arkiven – befintliga och kommande – behöver överföras till digital form (digitiseras), likaså behöver det ökade inflödet av digitala arkiv tillgängliggöras. Digitisering underlättar inte bara åtkomst till källorna, utan möjliggör också utveckling inom AI. Riksarkivet hanterar tillsammans med andra statliga myndigheter en mycket stor mängd information som kan ligga till grund för svenska AI-modeller utvecklade med hjälp av öppna, representativa och kontrollerade källor av hög kvalitet. Riksarkivet ligger i framkant vad gäller utveckling av AI-modeller som stöder datadriven forskning och kan också samla arkiv från hela arkivfären i den Nationella Arkivdatabasen (NAD).

Riksarkivets förslag

I det följande beskrivs och motiveras tre förslag som syftar till att möjliggöra forskning av hög kvalitet.

- 1) Tillägg i förordning 2009:1593 med instruktion för Riksarkivet om att utgöra en forskningsinfrastruktur.

I takt med att forskningen i allt högre grad baseras på forsknings- och e-infrastrukturer bör det klargöras i myndighetens instruktion att Riksarkivet utgör och har till uppgift att tillhandahålla en sådan.

Riksarkivet som organisation är väl förberedd på att möta den snabba utvecklingen inom AI och forskningens behov. Vid myndigheten finns disputerade forskare och en ny avdelning för *Forskning och kunskapsuppbyggnad* har inrättats under året. Inom denna samlas flera verksamheter och kompetenser. Bland annat utvecklar och tillgängliggör Riksarkivet AI-modeller för maskinell tolkning av handskriven och tryckt text samt språkmodeller. Modellerna är utvecklade med hjälp av öppen källkod och tillgängliggörs under öppen licens vilket innebär att de är helt fria att använda eller modifiera för eget bruk.⁴ Riksarkivet deltar i READ-COOP som är en European Cooperative Society som utvecklar den AI-baserade plattformen Transkribus.⁵ Inom myndighetens särskilda uppdrag för Forskning och utveckling (FoU) prioriteras deltagande i tre av Vetenskapsrådet understödda nationella forskningsinfrastrukturer:

³ SOU 2021:65, s. 23, 90, 151, 209–210, 317, 327.

⁴ <https://huggingface.co/Riksarkivet>

⁵ <https://readcoop.eu/>

SwedPop, Språkbanken och Huminfra.⁶ Riksarkivet bidrar inom dessa med datakurering, metod- och teknikutveckling i syfte att främja ny forskning. Riksarkivet bedriver också FoU-projekt i syfte att kombinera medborgarforskning med AI, vilket både möjliggör och förbättrar våra AI-modeller, samt bidrar till medskapande och ökad transparens.

Riksarkivet har också en särskild enhet för storskalig digitisering.⁷ Denna digitiserar Riksarkivets egna samlingar, samt erbjuder digitisering som tjänst till myndigheter och offentliga verksamheter.

Riksarkivets söktjänster omfattar, förutom Sök i NAD, även Digitala forskarsalen med ett 30-tal specialiserade forskningsdatabaser och register.⁸ Riksarkivet tillhandahåller en storskalig åtkomst till data genom publika API:er och nedladdningsbara dataset i vilka datainnehållet är fritt för vidareutnyttjande (öppna data).⁹ Riksarkivets öppna data tillgängliggörs bland annat i de nationella forskningsinfrastrukturerna.

Riksarkivet bedriver också vetenskaplig utgivningsverksamhet genom redaktioner för Svenskt Diplomatarium, Svenskt biografiskt lexikon och Det medeltida Sverige, som också tillhandahåller egna söktjänster ämnade för forskare.

2) Nationell Arkivdatabas som e-infrastruktur för forskning.

För att kunna bedriva forskning med hög kvalitet och effektivitet behöver forskarna tillgång till forsknings- och e-infrastrukturer. NAD är nyckeln till information om arkiv i Riksarkivet, kommun- och regionarkiv, folkrörelse- och föreningsarkiv, näringslivsarkiv, museer, bibliotek med flera. NAD är således en oumbärlig resurs för forskare som använder arkivmaterial. De senaste åren har stora satsningar skett på uppbyggande av nya nationella forskningsinfrastrukturer.¹⁰ Det behövs emellertid även satsningar som inriktar sig på att utveckla de existerande, såsom NAD.

För att främja datadriven forskning och innovation behöver Riksarkivets digitala infrastruktur utvecklas. NAD behöver utvecklas för att kunna ta emot, lagra och tillgängliggöra den förväntade stora mängd information om arkiven som skapas med hjälp av AI-modeller och som utvecklas i samverkan med forskarsamhället och andra aktörer i samhället.

Riksarkivet har, i samverkan med bland annat Riksarkivet i Finland, de senaste åren utvecklat AI-modeller som med hög precision tolkar handskrivna text. Miljontals handskrivna dokument i arkiven, som är tillgängliga som skannade bilder, kan därmed bli tillgängliga och sökbara på innehållsmässig nivå. Riksarkivet har även utvecklat språkmodeller som bland annat kan användas för entitetsigenkänning, semantisk sökning och topic modeling. AI-modeller av detta slag kan vidare användas för att

⁶ <https://swedpop.se/>; <https://www.sprakbanken.se/>; <https://www.huminfra.se/>

⁷ <https://riksarkivet.se/digitisering>

⁸ <https://sok.riksarkivet.se/digitala-forskarsalen>

⁹ <https://sok.riksarkivet.se/data-api>

¹⁰ <https://www.vr.se/uppdrag/forskningsinfrastruktur.html>

automatisera processerna för hur arkivinformaton sammanställs, registreras och kontrolleras. Denna utveckling öppnar för helt ny forskning som analyserar arkiven på ett sätt som tidigare inte varit möjligt. Den effektiviserar också verksamheten då manuellt arbete kan automatiseras. Till exempel har Riksarkivet med hjälp av egenutvecklade AI-modeller indexerat 8 miljoner sidor från fastighetsböcker. Forskning och utveckling inom AI-området ger med andra ord stor forskningsnytta och effektiviserar och förbättrar verksamheten i sig. Indexering och annat tillgängliggörande som skulle ta årtal att utföra manuellt, och som därför inte är möjligt att genomföra, kan nu genomföras med ny teknik. För att tillvarata de möjligheter AI innebär och för att utveckla NAD har Riksarkivet behov av ett årligt ökat anslag om 15 Mkr.

3) Storskalig digitisering av samlingar i Riksarkivet inklusive infrastruktur och beräkningskraft för utveckling av AI-modeller.

Ett grundläggande problem är att Riksarkivets nuvarande anslag inte räcker till för den storskaliga digitisering som behövs för att tillgodose forskningens behov.¹¹ Idag digitiseras arkivhandlingar för forskningsändamål primärt efter extern förfrågan genom beställningar från enskilda forskare eller forskningsprojekt. Denna möjlighet är viktig i sig, eftersom forskning då kan bedrivas platsoberoende. Dessutom bevaras och skyddas originalhandlingar. Men det ger inte förutsättning för storskalig digitisering eller forskning om hela samlingar och arkiv. De särskilda satsningar som gjorts på att digitisera kulturarvssamlingar, såsom forskningsprogrammet DIGARV, har inte haft någon effekt på behovet av digitisering för forskningsändamål inom Riksarkivet.¹²

Riksarkivet menar att storskalig digitisering måste planeras och ske på ett samordnat sätt – inte spridas ut på enskilda forskningsprojekt. Forskarna har en viktig roll för att initiera och lyfta vilka behov som finns, men det är Riksarkivet som behöver genomföra digitiseringen inklusive bedöma behov av bevarandetekniska åtgärder och annan preparering. Riksarkivet välkomnar forskningsprojekt som inkluderar digitisering, men det blir ofta för dyrt att inkludera hela, obrutna arkiv och samlingar inom ramen för projekt. Enskilda forskare kan inte åläggas ansvar för att digitisera vårt gemensamma kulturarv. Istället borde en finansieringsmodell utarbetas som tillvaratar forskningens behov och Riksarkivets möjligheter att på bred front digitisera och långsiktigt säkra viktig samhällsinformation.

För att säkerställa tillgång till ett brett urval av arkiv, för att kunna dra nytta av digitala metoder och för att främja datadriven forskning måste alltså en storskalig satsning genomföras på att digitisera arkiv och samlingar. Riksarkivet förvarar 75 hyllmil analoga arkivhandlingar och har ett kontinuerligt inflöde av både analogt och digital material. Riksarkivet bedömer att fem procent av de analoga samlingarna är digitalt tillgängliga vilket innebär över 230 miljoner bilder med tillhörande metadata.

¹¹ Se SOU 2019:58, s 318–323, 547–548

¹² <https://www.digarv.se/>

Riksarkivets digitala arkiv utgör med andra ord en oumbärlig resurs, men om inte digitaliseringstakten ökar finns en risk att forskningen framöver begränsas till de arkiv som redan är digitalt tillgängliga.

Exempelvis är de centrala arkiven över kanslierna och statsdepartementen från 1500-talet till och med 1900-talet inte digitiserade. Dessa omfattar omkring 6 000 volymer eller 9 miljoner sidor text (645 hyllmeter). Att digitisera och skapa metadata för dessa skulle uppskattningsvis kosta ca 20 miljoner kronor. Om det sker blir det möjligt att forska kring den centrala politiska beslutskedjan från medeltiden och framåt, vilket skulle revolutionera möjligheten att bedriva forskning om svensk historia och samhällets politiska utveckling.

Storskalig digitisering av arkiv ger en möjlighet för Sverige som nation att utveckla AI-modeller som är öppna, representativa och bygger på kontrollerade källor av hög kvalitet. Arbete med och utveckling av dessa modeller förutsätter tillgång till hårdvara med betydande beräkningskraft. För vissa projekt har Riksarkivet haft tillgång till EU-kommissionens initiativ för superdatorer, EuroHPC JU.¹³ Riksarkivet behöver emellertid även tillgång till lokal beräkningskraft, om än i mindre skala. Brist på beräkningskraft är en betydande risk som kan innebära att den fortsatta utvecklingen och implementeringen av AI-modeller i verksamheten förhindras eller stannar av. En samordnad organisering behövs, då flera myndigheter står inför liknande utmaningar.

Mot denna bakgrund föreslås en satsning på digitisering av samlingar i Riksarkivet. Riksarkivet ställer sig därmed bakom Vetenskapsrådets rekommendation om att en storskalig satsning bör genomföras på digitalisering av bibliotekens, museernas och arkivens samlingar för att säkerställa att de bevaras och görs tillgängliga.¹⁴ Riksarkivet har både kapaciteten och förutsättningarna att påbörja ett sådant uppdrag med kort varsel. Riksarkivet behöver ett årligt ökat anslag om 30 miljoner kronor för digitisering och investeringar i beräkningskraft för AI-modeller. Riksarkivet menar att en satsning på storskalig digitisering och utbyggd infrastruktur vid myndigheten är det mest effektiva sättet att säkra forskares tillgång till arkiv i digital form.

Beslut i detta ärende har fattats av riksarkivarien Karin Åström Iko. Ärendet har handlagts av forskningsledare Olof Karsvall, föredragande. Vid ärendets slutliga beredning har avdelningschef Camilla Niss deltagit.

Karin Åström Iko

Olof Karsvall

¹³ <https://eurohpc-ju.europa.eu/>

¹⁴ Vetenskapsrådet, Forskningsöversikt 2023: Humaniora och samhällsvetenskap, s 5, 30.