

Inriktningsförslag för organisering av svensk e-infrastruktur för forskning

**Framtaget av Vetenskapsrådet och
Sveriges universitets- och högskoleförbund**

Inriktningsförslag för organisering av svensk e-infrastruktur för forskning

Dnr 4.5-2017-06624
ISBN 978-91-88943-27-9

Swedish Research Council
Vetenskapsrådet
Box 1035
SE-101 38 Stockholm, Sweden

Innehållsförteckning

Förord.....	3
Sammanfattning	4
1. Inledning.....	6
1.1 Läsanvisningar och disposition.....	7
2. Mot en ny sammanhållen organisation för forskningens e-infrastruktur	8
2.1 Organisationens struktur.....	9
3. Nuvarande organisation av svensk e-infrastruktur.....	12
3.1 Swedish National Infrastructure for Computing – SNIC.....	12
3.2 Svenska universitetsdatanätverket – Sunet.....	13
3.3 Svensk nationell datatjänst – SND.....	14
3.4 Vetenskapsrådets registeruppdrag.....	14
3.5 Microdata online access – MONA.....	15
3.6 Omfattande men fragmenterat	16
4. Vad behöver lösas?.....	18
4.1 Hantering av forskningsdata under hela forskningsdatalivscykeln	18
4.1.1 Forschargenererade data.....	21
4.1.2 Registerdata	23
4.1.3 Hälsa- och vårddata.....	24
4.1.4 Övriga data	25
4.2 Tillhandahållande av nät-, lagrings- och beräkningsresurser.....	25
4.2.1 Nät.....	25
4.2.2 Lagring	25
4.2.3 Beräkning.....	26
4.3 Internationella samarbeten.....	26
5. En något mer detaljerad beskrivning av inriktningsförslaget	28
5.1.1 Hantering av data	28
5.1.2 Nät-, lagrings- och beräkningsresurser.....	33
5.1.3 Övriga tjänster.....	33
5.1.4 Nationell strategi och internationella samarbeten	34
5.1.5 Organisationsform och ledningsstruktur.....	34
5.1.6 Finansiering	35
6. Lista över förkortningar och termer.....	37

Förord

Vetenskapsrådet och Sveriges universitets- och högskoleförbund (SUHF) presenterar här ett gemensamt framtaget inriktningsförslag som syftar till en genomgripande omorganisering av Sveriges nationella e-infrastrukturer för forskning, det vill säga de verksamheter som berör beräkning, lagring och överföring av data samt tillgängliggörande av data för forskning.

Vetenskapsrådet och Universitetens referensgrupp för forskningsinfrastruktur (URFI) påbörjade år 2017 en gemensam översyn av nationella e-infrastrukturer för forskning. Arbetet genomfördes av en expertgrupp som i slutet av 2018 presenterade rapporten *An outlook for the national roadmap for e-infrastructure for research*. I rapporten beskrivs det svenska e-infrastrukturlandskapet som fragmenterat vad gäller både styrning och finansieringsströmmar. Vidare pekar rapporten på att bristen på en sammanhållen nationell strategi försvårar både utvecklingen av en effektiv nationell e-infrastruktur och Sveriges engagemang i de allt mer omfattande internationella initiativen inom området.

Under 2019 tillsatte Vetenskapsrådet och SUHF en gemensam arbetsgrupp med uppdrag att ta expertgruppens arbete vidare och formulera ett inriktningsförslag för en sammanhållen organisering av Sveriges nationella e-infrastrukturer för forskning. Arbetsgruppen bestod av fyra företrädare för Vetenskapsrådet och fyra företrädare för SUHF, de senare även ledamöter i URFI:

Lars Börjesson, (Ledamot i URFI, CTH)
 Per Dannetun (Ledamot i URFI, LiU)
 Björn Halleröd (Huvudsekreterare för RFI)
 Lars Kloo (Huvudsekreterare för NT)
 Hjalmar Laudon (Ledamot i RFI, SLU)
 Ingela Nyström (Ledamot i RFI, UU)
 Katrine Riklund (Ordf. i URFI, UmU)
 Stacey Sörensen (Ledamot i URFI, LU)

I sitt arbete har gruppen haft tjänstemannastöd av Monica Lassi vid Lunds universitet och Hanifeh Khayyeri vid Vetenskapsrådet.

Både Vetenskapsrådets och SUHF:s styrelser har ställt sig bakom den övergripande inriktningen som syftar till en sammanhållen organisation för svensk e-infrastruktur för forskning, vilket båda parter ser som ett viktigt steg för att stärka svensk forskning. Det är samtidigt viktigt att, med detta övergripande mål i sikte, betona behovet av fortsatt utredning av de exakta formerna för hur detta ska förverkligas.

Stockholm, 6 februari 2020

Sven Stafström
 Generaldirektör, Vetenskapsrådet

Astrid Söderbergh Widding
 Ordförande, SUHF

Sammanfattning

I denna rapport formulerar Vetenskapsrådet tillsammans med Sveriges universitets- och högskoleförbund ett inriktningsförslag för en ny organisation av svensk e-infrastruktur för forskning. Att utveckla forskningens digitala verktyg är nödvändigt för att garantera forskningens kvalitet och förhindra att bristande tillgång till e-infrastruktur utvecklas till en flaskhals för betydande delar av forskningssystemet. Utgångspunkten för rapporten är den översyn som Vetenskapsrådet tillsammans med Universitetens referensgrupp för forskningsinfrastruktur (URFI) presenterade våren 2019 (*An outlook for the national roadmap for e-infrastructure for research*). Översynen beskriver ett fragmenterat landskap, bristande samordning och avsaknad av nationell strategi.

Det är angeläget att skapa förutsättningar för Sverige att utveckla en e-infrastruktur för forskning som möter framtidens behov, tar till vara på de möjligheter Sverige har som forskningsnation och medverkar till att göra forskningsdata öppet tillgänglig. För att åstadkomma detta föreslår vi en ny organisation för forskningens e-infrastruktur. I rapporten har denna organisation getts arbetsnamnet NYe-Infra. Inom ramen för NYe-Infra vill vi sammanföra fem existerande verksamheter: Swedish National Infrastructure for Computing (SNIC), Svenska universitetsnätverket (SUNET), Svensk nationell datatjänst (SND), Vetenskapsrådets registeruppdrag och Statistiska centralbyråns system för att ge forskare tillgång till forskningsdata som innehåller personuppgifter (MONA). *Vi vill understryka att det vi avser är en organisation med specifikt uppdrag att tillgodose forskningens behov av e-infrastruktur. Administrativa system som berör forskning och högre utbildning eller annan offentlig verksamhet berörs inte av inriktningsförslaget.*

Inriktningsförslaget grundar sig på forskningens behov. Övergripande uppgifter som bör ingå i NYe-Infra är att:

- utarbeta och implementera en sammanhängande strategi för att tillgodose svensk forsknings behov av e-infrastruktur.
- verka för att svensk e-infrastruktur för forskning har långsiktigt förutsägbara ekonomiska förutsättningar.
- på ett samlat sätt företräda Sverige i det allt mer omfattande internationella e-infrastruktursamarbetet.
- se till att den tekniska utvecklingen av e-infrastruktur sker i samklang med omvärlden, är öppen och interoperabel, och samtidigt tillgodoser krav på säkerhet och integritet.
- se till att policyer och strategier på forskningsdata- och e-infrastrukturuområdena är synkroniserade med andra intressenter nationellt och internationellt.
- tydliggöra förutsättningarna för och ge stöd till individuella forskare forskningsdatastödfunktioner på lärosätena vad gäller rättigheter och skyldigheter för tillgängliggörande av forskningsdata.
- bidra till att svensk forskning kan dra nytta av tillgången till unika registerdata och data från hälso- och sjukvården samtidigt som skyddet för den personliga integriteten stärks.

- synkronisera med andra myndigheter med beröringspunkter, exempelvis Etikprövningsmyndigheten, DIGG och Riksarkivet.

Det förslag på ny organisation för forskningens e-infrastruktur som presenteras ska ses som ett förslag på inriktning och en grund för fortsatt utredning. Det finns en rad både praktiska och legala aspekter som måste utredas mer i detalj innan NYe-Infra kan realiseras. Vår förhoppning är att denna rapport kan användas som underlag och tydlig inriktning för en utredning av hur organisationen av svensk e-infrastruktur för forskning ska se ut i framtiden.

1. Inledning

I denna rapport presenterar Vetenskapsrådet och Sveriges universitets- och högskoleförbund (SUHF) ett gemensamt inriktningsförslag rörande framtida organisering av svensk e-infrastruktur för forskning. Med e-infrastruktur för forskning avses resurser för lagring, överföring, tillgängliggörande, beräkning och analys av digitala data. Inom begreppet ryms också de kompetenser och organisationer som är nödvändiga för att forskningen ska kunna tillgodogöra sig befintliga resurser och existerande data. Vi vill understryka att rapporten innehåller ett inriktningsförslag. Det betyder att vi förutsätter att en mer grundläggande utredning initieras innan en eventuell omorganisering av svensk e-infrastruktur för forskning eventuellt genomförs.

E-infrastruktur spelar en allt viktigare roll för Sveriges möjlighet att bibehålla sin position som en ledande forskningsnation. Ökade krav på beräkningskapacitet, utveckling inom maskininlärning och artificiell intelligens (AI), allt större och mer komplexa datamängder driver utvecklingen. Kraven accentueras av strävan att göra forskningsdata öppna och återanvändningsbara. Ur såväl internationellt som nationellt perspektiv ökar forskningens behov av avancerad e-infrastruktur. Utvecklingen rör alla områden, allt ifrån astrofysik till humaniora. Forskningens digitalisering innebär också att behoven av support och vägledning ökar inom flera områden. Det är en utveckling som sedan länge funnits inom medicinsk forskning men som nu också är tydlig inom samhällsvetenskap och humaniora. En konsekvens av utvecklingen är att behoven av internationella samarbeten ökar, vilket medför att Sverige behöver en sammanhållen strategi och förmåga att vara en tydlig partner.

En ny organisation för e-infrastruktur för svensk forskning bör ha som övergripande mål att:

- säkra forskningens kvalitet genom att säkerställa hög kvalitet i Sveriges e-infrastruktur och forskningsdatahantering.
- säkerställa Sveriges plats i frontlinjen av den internationella forskningen genom att utveckla Sveriges e-infrastruktur på ett väl sammanhållet sätt.
- skapa förutsättningar för att Sveriges e-infrastrukturresurser nyttjas på ett effektivt sätt, och därmed kapa ledtider från forskningsidé till publicering.
- skapa möjligheter till nya forskningsfrågor och metoder, bland annat genom att möjliggöra att enkelt kunna kombinera datamängder och datakällor på nya sätt och genom samverkan med relevanta aktörer för att underlätta forskningsprocessen.
- tillhandahålla en samlad strategi för Sverige på nationell och internationell nivå gällande frågor om forskningsdata och e-infrastruktur. Detta inkluderar att ta fram policy inom relevanta områden samt att driva policyfrågor nationellt och internationellt för synkronisering med bland annat European Open Science Cloud (EOSC) och The European High-Performance Computing Joint Undertaking (EuroHPC).

Behovet av att formulera ett inriktningsförslag rörande Sveriges e-infrastruktur för forskning emanerar från den översyn av området som Vetenskapsrådet tillsammans

med universitetens referensgrupp för forskningsinfrastruktur (URFI) initierade år 2018. Översynen genomfördes av en expertpanel och arbetet presenterades våren 2019 i rapporten *An outlook for the national roadmap for e-infrastructure for research*¹. Rapporten pekar på ett fragmenterat landskap, bristande samordning och avsaknad av nationell strategi. Panelen menar att detta gör det svårare för Sverige och svenska forskare att delta i den snabba omvandlingen som sker i och med allt större behov av digital forskningsinfrastruktur och arbetet med öppna forskningsdata. Samtidigt bör det framhållas att enskilda komponenter i landskapet är väl utvecklade och att det finns en betydande potential att ytterligare utveckla svenska e-infrastruktur för forskning. Efter att rapporten publicerats bildade Vetenskapsrådet och URFI en informell arbetsgrupp för att ta rapportens slutsatser vidare och arbeta fram ett inriktningsförslag rörande framtida organisering av svenska e-infrastruktur. Gruppen har därefter fått ett gemensamt mandat från Vetenskapsrådet och SUHF. I uppdragsbeskrivningen till gruppen står bland annat att:

Vetenskapsrådet och SUHF är överens om att frågan om en översyn av svensk e-infrastruktur är av hög prioritet. Vetenskapsrådet och SUHF tillsätter därför en gemensam arbetsgrupp med uppdrag att föreslå en modell för hur Sveriges nationella e-infrastrukturer för forskning ska organiseras. Förslaget på modell ska innefatta såväl ägarskap som finansiering och ska beakta både utnyttjandet och driften av infrastrukturerna. Målet med förslaget ska vara att skapa en modell som ger högsta möjliga effektivitet för det svenska forskningssystemet.

1.1 Läsanvisningar och disposition

I det kommande avsnittet presenterar vi översiktligt vårt förslag om en ny sammanhållen e-infrastruktur för svensk forskning. Syftet är att här peka ut en övergripande inriktning. Information om inriktningsförslagets innehåll i detalj finns i rapportens avslutande kapitel 5, där vi också diskuterar frågor som rör ledning och styrning samt finansiering. Tyngdpunkten ligger dock på vilka fördelar vi ser med att samordna de nu existerande e-infrastrukturverksamheterna. För att sätta utredningen i sitt sammanhang beskriver vi i kapitel 3 kortfattat det nuvarande svenska e-infrastrukturlandskapet och de fem verksamheter som mest direkt berörs av inriktningsförslaget. I kapitel 4 beskriver vi de konkreta frågor rörande e-infrastruktur för forskning som vi anser vara mest angelägna att åtgärda.

¹ <https://www.vr.se/analys/rapporter/vara-rapporter/2019-02-14-an-outlook-for-the-national-roadmap-for-e-infrastructures-for-research.html>

2. Mot en ny sammanhållen organisation för forskningens e-infrastruktur

Syftet med inriktningsförslaget är att adressera en rad utmaningar som svensk e-infrastruktur för forskning står inför. Att utveckla de digitala verktygen är nödvändigt för att garantera forskningens kvalitet och förhindra att bristande tillgång till e-infrastruktur utvecklas till en flaskhals för betydande delar av forskningssystemet. Samtidigt krävs det både långsiktiga och övergripande strategiska överväganden och långsiktigt förutsägbara finansiella förutsättningar.

För att möta forskningens behov av e-infrastruktur, kraven på öppna forskningsdata och ökad tillgänglighet till registerdata behöver Sverige en ny sammanhållen organisation för forskningens e-infrastruktur, fortsättningsvis benämnd **NYe-Infra**.

Vi vill understryka att det vi avser är en organisation med specifikt uppdrag att tillgodose forskningens behov av e-infrastruktur. Administrativa system som berör forskning och högre utbildning, exempelvis Ladok, Prisma och NyA, eller annan offentlig verksamhet berörs inte av inriktningsförslaget. Vi utesluter dock inte att ansvar för denna typ av system skulle kunna tillföras NYe-Infra, men det bör i så fall ske först i ett senare skede efter det att NYe-Infra etablerats. Förslaget kommer även att ha betydelse för högre utbildning då mycket av det som berörs även har relevans för denna. Fokus i detta dokument kommer dock vara forskningens behov och vårt grundantagande är att NYe-Infra även kommer att förbättra förutsättningarna för högre utbildning. Myndigheten för digital förvaltning (DIGG) startade sin verksamhet år 2018 och har som uppgift att samordna och stödja den förvaltningsgemensamma digitaliseringen i syfte att göra den offentliga förvaltningen mer effektiv och ändamålsenlig. Ett framtida NYe-Infra förväntas samarbeta med DIGG men vill vi understryka att den verksamhet som bedrivs av DIGG i huvudsak inte överlappar med NYe-Infra.

Inriktningsförslaget har betydelse för stora delar av det svenska forskningssystemet. Den organisationsförändring som diskuteras är dock begränsad till i huvudsak fem existerande verksamheter: Swedish National Infrastructure for Computing (SNIC), Svenska universitetsnätverket (Sunet), Svensk nationell datatjänst (SND), Vetenskapsrådets registeruppdrag och Statistiska centralbyråns system för att ge forskare tillgång till forskningsdata som innehåller personuppgifter (MONA). Verksamheterna beskrivs närmare i kapitel 3. Vi menar att dessa fem verksamheter bör slås samman till en ny organisation: NYe-Infra. För att en sådan organisation ska vara motiverad krävs det att den substantiellt bidrar till att förbättra förutsättningarna för svensk forskning vad gäller hantering av forskningsdata, tillgodose behovet av beräkningskapacitet och internationella samarbeten.

Det förslag på en ny organisation för forskningens e-infrastruktur som presenteras här är ska som nämnts ovan ses som ett förslag på inriktning och en grund för fortsatt utredning. Det finns en rad praktiska och legala aspekter som måste utredas mer i detalj innan NYe-Infra kan realiseras. En central komponent i vårt förslag är att NYe-Infra ska kunna förvalta och hantera data från myndighetsregister och data från sjukvården som innehåller personuppgifter och om relevant annan data belagd

med sekretess. Detta ska dock endast gälla för de specifika data som, efter gängse prövning, lämnats ut för forskning (se avsnitt 5.1.1). Vi ser det som en central uppgift för en vidare utredning att klarlägga de legala förutsättningarna för detta och om nödvändigt föreslå en lagstiftning som gör en sådan hantering möjlig. Förslaget berör hantering av andra myndigheters och sjukvårdens data och vilar på att införandet av NYe-Infra kan genomföras i nära samarbete med dessa organisationer. I vissa fall, speciellt gäller detta Sunet, finns det avnämare utanför forskningsystemet. Deras behov måste även fortsättningsvis kunna tillgodoses. De nuvarande organisationerna SNIC, Sunet, SND, Vetenskapsrådets registeruppdrag och MONA måste få möjlighet att lämna synpunkter, något som kommer att vara av stort värde.

Forskarnas och forskningens behov bör vara utgångspunkten för NYe-Infras verksamhet. Övergripande uppgifter som bör ingå i NYe-Infra är att:

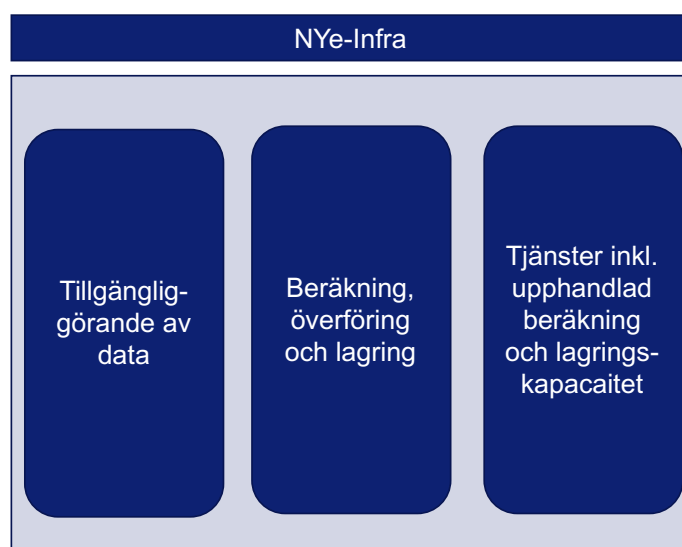
- utarbeta och implementera en sammanhängande strategi för att tillgodose svensk forsknings behov av e-infrastruktur.
- verka för att svensk e-infrastruktur för forskning har långsiktigt förutsägbara ekonomiska förutsättningar.
- på ett samlat sätt företräda Sverige i det allt mer omfattande internationella e-infrastruktursamarbetet.
- se till att den tekniska utvecklingen av e-infrastruktur sker i samklang med omvärlden, är öppen och interoperabel, och samtidigt tillgodoser krav på säkerhet och integritet.
- se till att policyer och strategier på forskningsdata- och e-infrastruktururområdena är synkroniserade med andra intressenter nationellt och internationellt.
- tydliggöra förutsättningarna för och ge stöd till individuella forskare forskningsdatastödfunktioner på lärosätena vad gäller rättigheter och skyldigheter för tillgängliggörande av forskningsdata.
- bidra till att svensk forskning kan dra nytta av tillgången till unika registerdata och data från hälso- och sjukvården samtidigt som skyddet för den personliga integriteten stärks.
- synkronisera med andra myndigheter med beröringspunkter, exempelvis Etikprövningsmyndigheten, DIGG och Riksarkivet.

2.1 Organisationens struktur

Tentativt tänker vi oss att NYe-Infra organiseras i tre avdelningar (se figur 1). Den första innefattar funktioner för att stödja dokumentation och tillgängliggörande av forskningsdata. Av de existerande verksamheterna täcker avdelningen främst de verksamheter som idag hanteras av SND och Vetenskapsrådets registeruppdrag. Båda dessa verksamheter verkar för att forskningsdata dokumenteras enligt givna standarder, att metadata sammanställs och används för att göra uppgifter om existerande datakällor sökbara. Det pågår redan idag ett erfarenhetsutbyte mellan SND och registeruppdraget vad gäller sökverktyg och interoperabilitet mellan de verktyg som var och en nu utvecklar. Deras uppgifter är till viss del överlappande när det gäller forskargenererade data som innehåller personuppgifter och där individer är identifierbara via personnummer. I huvudsak vänder sig dock SND och

registeruppdraget mot olika typer av datakällor. SND har ansvar för att tillgängliggöra forskargenererade data, registeruppdraget har som syfte att öka tillgängligheten till myndigheters registeruppgifter och sjukvårdens data. Vi ser att det potentiellt finns betydande synergieffekter i att sammanföra de nuvarande verksamheterna och att en kommande utredning bör beakta dessa. Vi menar också att NYe-Infra på ett mer koordinerat sätt kan driva arbetet med öppna forskningsdata och Sveriges engagemang i EOSC.

Figur 1. Övergripande tentativ organisering av NYe-Infra.



Den andra avdelningen utgörs av det som kan beskrivas som den svenska forskningens nationella e-infrastruktur, inklusive hårdvara, tjänster, systemnära expertis och användarnära expertstöd. Av de existerande verksamheterna är det SNIC, Sunet och MONA som främst omfattas. Att på så sätt skapa en sammanhållen nationell verksamhet för överföring, lagring och beräkning skulle öka Sveriges möjligheter att svara upp mot forskningens krav och internationella samarbeten.

Den tredje avdelningen organiserar olika typer av tjänster. Framförallt Sunet erbjuder redan idag lärosäten och övriga avnämare en rad tjänster vid sidan av själva nätverket. En viktig del av dessa rör lagring av forskningsdata. Vi ser framför oss att betydelsen av tjänster kommer att öka. I vissa fall krävs att kapacitet för vissa tjänster byggs upp internt av NYe-Infra. Exempelvis gäller det kapacitet att på uppdrag av lärosäten lagra data som innehåller personuppgifter. I andra fall kan NYe-Infra upphandla kommersiella tjänster som sedan erbjuds avnämarna. Det kan handla om lagring av forskningsdata som inte innehåller personuppgifter, beräkningskapacitet för att exempelvis klara tillfälliga belastningstoppar, och en rad andra tjänster. Behoven av denna typ av tjänster kommer förmodligen att öka och bli en viktig del av NYe-Infra. Vi menar också att tjänster måste hanteras på ett transparent och tydligt sätt. En eventuell utredning bör särskilt beakta under vilka förutsättningar NYe-Infra kan erbjuda tjänster som upphandlas av kommersiella aktörer.

Att skapa en ny organisation, ett NYe-Infra för svensk forskning, kan inte motiveras enbart utifrån de eventuella synergieffekter som kan uppstå genom att

endast omorganisera existerande verksamheter. Redan idag finns ett samarbete mellan de ingående verksamheterna. Att ytterligare utveckla detta är ett sätt att möta utmaningarna inom svensk e-infrastruktur för forskning. Vi menar dock att en ny organisation kan stärka svensk forskning på ett sätt som inte är möjligt inom ramen för nuvarande organisering. Vid sidan av de övergripande uppgifter som listas ovan gäller det exempelvis hantering av forskningsdata under hela forskningsdatalivscykeln, tillgängliggörande av data samt lagring och analys av data som innehåller känsliga personuppgifter. I kapitel 5 diskuteras det som vi ser som centrala vinster med att inrätta NYe-Infra mer i detalj.

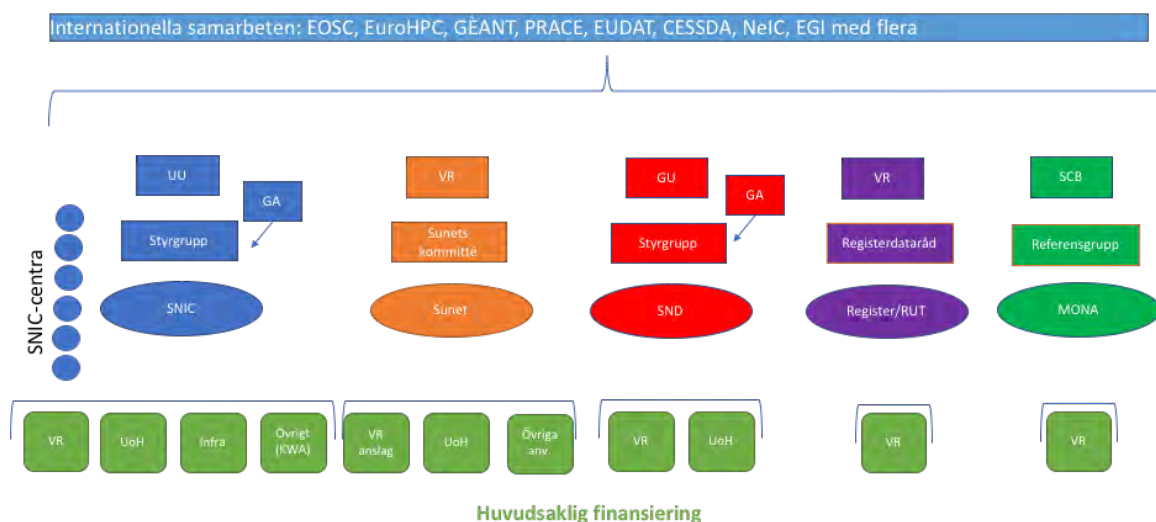
För att ta vara på de erfarenheter och kompetenser som finns i e-infrastrukturlandskapet, bör NYe-Infra vara en distribuerad organisation som tillåter att medarbetare finns utspridda i landet. De existerande verksamheter som vi föreslår ska ingå i NYe-Infra är redan idag distribuerade mellan olika lärosäten. Detta stärker förutsättningarna för NYe-Infra att bygga starka relationer med såväl forskare som stödfunktioner och lärosäten.

Det viktigaste råd som framförs i *An outlook for the national roadmap for e-infrastructure for research* är att be regeringen tillsätta en nationell kommitté med uppdrag att samordna de olika intressenterna. Vi håller med om att en bättre samordning behövs. Men en nationell kommitté som läggs ovanpå de existerande organisationerna med deras styrelseformer intakta, har små eller inga möjligheter att minska fragmenteringen av e-infrastrukturlandskapet. Vi menar tvärtom att en nationell kommitté utan ett tydligt ägaransvar riskerar att leda till att komplexiteten och otydligheten ökar. Vi anser därför att den kommitté som föreslås i rapporten *An outlook for the national roadmap for e-infrastructure for research* inte bör tillsättas. Istället bör, som föreslagits ovan, en ny sammanhållen organisation, som kan garantera organisatorisk, ekonomisk, juridisk och teknisk långsiktighet och hållbarhet, övervägas.

3. Nuvarande organisation av svensk e-infrastruktur

En av de viktigaste slutsatserna från den översyn av e-infrastrukturlandskapet som Vetenskapsrådet och URFI genomfört är att organiseringen kännetecknas av fragmentering. Figur 2 visar en övergripande bild av några av de mest centrala delarna av de svenska e-infrastrukturlandskapet och också dess komplexitet. För var och en av de fem berörda verksamheterna visar figuren översiktligt finansieringsströmmar, styrning och ägande.

Figur 2. Skiss över Sveriges e-infrastrukturlandskap



3.1 Swedish National Infrastructure for Computing – SNIC

SNIC är ett konsortium av tio universitet, med Uppsala universitet som värdunderstitet. SNIC:s stämma (General Assembly) utgörs av representanter från de universitet och högskolor som bidrar till finansieringen, och skall värna deras intressen. Styrelsen för SNIC tillsätts av Uppsala universitet efter samråd med SNIC:s stämma och Vetenskapsrådet. Centrala styrdokument är konsortialöverenskommelsen för SNIC och Vetenskapsrådets särskilda bidragsvillkor. Den fysiska infrastrukturen, dvs. SNIC:s beräknings- och lagringsresurser, är i dagsläget fördelad på sex olika lärosäten där det finns datorcentrum.² Respektive lärosäte har det formella ansvaret för respektive center. Därutöver finns personal som tillhandahåller avancerat användarstöd hos medlemmarna i konsortiet.

² Lunds Universitet, Chalmers tekniska högskola, Linköpings universitet, Kungliga tekniska högskolan, Uppsala universitet och Umeå universitet.

SNIC:s grundfinansiering kommer till lika delar från Vetenskapsrådet och de universitet och högskolor som utgör SNIC-konsortiet. Stora forskningsinfrastrukturer som nyttjar SNIC ska själva stå för sina kostnader. Vid sidan av det kan SNIC få bidrag från andra externa finansiärer. Tidigare har SNIC mottagit bidrag från Knut och Alice Wallenbergs Stiftelse (KAW) för att bygga upp kapacitet att hantera forskningsdata som innehåller känsliga persondata. Ett möjligt framtida externt bidrag gäller SNIC:s eventuella medverkan i de satsningar som nu görs på artificiell intelligens. SNIC ingår också i en rad internationella samarbeten indikerade i de gråa bubblorna till höger i figur 2. SNIC:s uppdrag är avgränsat till att förse forskning vid svenska lärosäten med nödvändiga beräkningsresurser, och därtill kopplad lagring och avancerat användarstöd, av en storlek som inte går att tillhandahålla lokalt vid enskilda lärosäten. Uppdraget innefattar även att delta i internationella samarbeten. SNIC bidrar idag till lagring av aktiv forskningsdata och kan bidra med långsiktiga resurser för lagring av forskningsdata.

Översiktlig fördelning av intäkter

Vetenskapsrådet	100 MKR
Universitet och högskolor	100 MKR
Totalt	200 MKR

3.2 Svenska universitetsdatanätverket – Sunet

Sunet har ett brett uppdrag att förse en lång rad, i huvudsak statliga, organisationer med nätverksuppkoppling. Även om universitets- och högskolesektorn volymmässigt är dominerande användare har Sunet ett uppdrag som sträcker sig utanför forskningen. Totalt är idag ca 115 statliga myndigheter och organisationer uppkopplade på Sunets forskningsnät. Vid sidan av datakommunikationskapacitet erbjuder Sunet olika typer av tjänster till sina användare. En allt viktigare del av tjänsterna är lagring av data. Sunet driver identitetsfederationen SWAMID som är en viktig del av Sunets tjänster. Kunskapen inom detta område har lett till ett flertal andra samarbeten, bland annat med DIGG och Skolverket. Även inom områdena it-säkerhet och robust datakommunikation sker samverkan med Post- och telestyrelsen (PTS), Myndigheten för samhällsskydd och beredskap (MSB) och andra relevanta parter i syfte att kunna hantera ökande behov inom flera samhällssektorer.

Sunet är en del av det globala forskningsnätet och samverkar därmed internationellt, bland annat genom delägarskap i NORDUnet och inom Europa via det EU-finansierade GÉANT-projektet där Sunet och övriga europeiska forskningsnätorganisationer är medlemmar.

Sunet är en avdelning på Vetenskapsrådet vilket gör att det yttersta ansvaret faller på Vetenskapsrådets styrelse. Sunet har även en kommitté som består av olika intressenter och som styr Sunet med ett mandat från Vetenskapsrådets generaldirektör. Sunet finansieras via ett riktat statligt anslag till Vetenskapsrådet samt av användaravgifter (nät och tjänster) från anslutna universitet och högskolor samt andra organisationer.

Översiktlig fördelning av intäkter (bokslut 2018)

Vetenskapsrådet anslag 3:9 Sunet	49 MKR
Avgifter (nät och tjänster)	166 MKR
Bidrag	6 MKR
Övriga intäkter	1 MKR
Totalt	222 MKR

3.3 Svensk nationell datatjänst – SND

SND arbetar för att tillgängliggöra forskningsdata och har under de senaste åren byggt upp en distribuerad verksamhet med övergripande syfte att öka tillgänglighet till forskningsdata och därmed bidra till målet om öppen vetenskap. Det arbetssätt som SND idag har anammats innebär att SND i huvudsak tillhandahåller information – metadata – om vilka datamängder som finns och hur dessa görs tillgängliga. Primärt lagrar eller arkiverar inte SND forskningsdata, även om det förekommer av historiska skäl. Den uppgiften faller istället på ägaren av datamängden, vilket vanligen är ett lärosäte. SND spelar en central roll i att åstadkomma ett system med öppna forskningsdata och är i den rollen en svensk resurs i relation till den europeiska satsningen på EOSC.

SND är en enhet vid Göteborgs universitet men har i likhet med SNIC både en styrgrupp och stämma. Finansieringen kommer från Vetenskapsrådet och de nio lärosäten som ingår i konsortiet bakom SND. Annan projektbaserad finansiering förekommer. En viktig del av SND:s verksamhet rör etablerandet av ett nätverk av forskningsdatastödfunktioner, inom SND kallat Data Access Units (DAU:er), vid de enskilda lärosätena. Syftet är att etablera en lokal forskningsdatastödfunktion med kompetens att stödja lokala forskare att göra forskningsdata tillgängliga och i förlängningen bistå medverkande organisationer i att försäkra sig om att forskningsdata kan återfinnas, förstås och återanvändas. I dagsläget medverkar 32 organisationer, de flesta är lärosäten men även RISE och Polarforskningssekreteriatet deltar. Uppbyggnaden, och den löpande verksamheten, finansieras av respektive medverkande organisation.

Översiktlig fördelning av intäkter

Vetenskapsrådet	14 MKR
Universitet och högskolor (inkl. konsortiets medverkan)	17 MKR
Övrigt (EU, CESSDA ERIC etc.)	4 MKR
Totalt	35 MKR

3.4 Vetenskapsrådets registeruppdrag

Registeruppdraget är ett uppdrag som regeringen gett till Vetenskapsrådet för att öka tillgängligheten till och därmed möjligheten att använda registerdata i forskning.

Registeruppdraget organiseras som en enhet vid Vetenskapsrådet och finansieras av ett särskilt anslag till Vetenskapsrådet. Kopplat till uppdraget finns också ett

registerdataråd som är sammansatt av representanter för de stora registerhållarna: Statistiska centralbyrån, Socialstyrelsen, Datainspektionen och biobanker samt forskarrepresentanter. Registerdatarådet är ett rådgivande organ.

Med registerdata avses de register som innehåller persondata – data med en personidentifikation, vanligen personnummer – som hålls av olika myndigheter och inom sjukvården, dvs. register som primärt inte har forskning som ändamål men som likväl är mycket värdefulla för forskning. Uppdraget hanterar även information som rör innehåll i biobanker med humana prover. Ett grundläggande syfte är att forskningen ska kunna utnyttja den forskningspotential som finns i de internationellt sett unika svenska registren. En viktig aspekt är att den forskning som använder sig av registeruppgifter i allt större grad bedrivs utanför lärosätena. Stora intressenter och användare av uppgifterna finns i hälso-och sjukvården och läkemedelsindustrin som använder samma registeruppgifter för att ta fram statistik för kvalitetsutveckling av vården och för uppföljning av läkemedel. Även flera myndigheter använder sig av uppgifterna för sitt arbete (t.ex. Läkemedelsverket, Tandvårds- och läkemedelsförmånsverket (TLV), Statskontoret). Forskning definieras således brett i registeruppdraget som har nära samarbetspartners även utanför universitet och högskola.

Översiktlig fördelning av intäkter

Vetenskapsrådet	50 MKR
Totalt	50 MKR

3.5 Microdata online access – MONA

Efter gängse etikprövning och prövning enligt Offentlighets- och sekretesslagen kan forskare få tillgång till mikrodata från register för ett specifikt forskningsändamål. MONA är SCB:s standardverktyg för att tillgängliggöra mikrodata efter godkänd prövning. För mikrodata som levereras av SCB erbjuder MONA en lösning där data lagras och analyseras utan att fysiskt lämna SCB. Uppgifterna är fortfarande utlämnade till en mottagare, men SCB agerar personuppgiftsbiträde genom att tillhandahålla plattformen. Om det i forskningsprojektet även behövs data från andra registerhållare kan dessa datamängder föras över till MONA. Den mottagande organisationen utser vilka personer som får logga in och arbeta med data i MONA. Arbetet sker via fjärruppkoppling i en windowsmiljö på SCB:s servrar och resultat i form av tabeller, grafer m.m. kan laddas ned till den egna datorn.

Med denna lösning behöver personuppgifter och andra mikrodata inte distribueras till forskaren och forskaren slipper bygga upp ett eget system för att garantera datasäkerhet. Därigenom bidrar MONA till att öka tryggheten och ge en enhetlig lösning för data som kan innehålla känsliga personuppgifter.

För närvarande nyttjar ca 1 500 användare, varav hälften forskare, den infrastruktur som MONA utgör och som omfattar hårdvara, mjukvara, säkerhet och lagring. MONA har funnits sedan 2005 och har hittills använts främst för registerforskning inom nationalekonomi, sociologi och andra samhällsvetenskapliga ämnen. Från och med 2019 har Socialstyrelsen godkänt att leveranser från deras register får föras över till MONA. På SCB pågår nu ett arbete för att öka

användandet av MONA när SCB lämnar ut data för medicinsk forskning där också registerdata från Socialstyrelsen ingår. MONA tillgängliggörs i nuläget endast för användare i de projekt där det ingår leveranser av mikrodata från SCB. MONA är alltså SCB:s leveransverktyg och ingen öppen plattform för bearbetning av data i andra projekt.

MONA administreras av SCB och till MONA finns en forskarsammansatt referensgrupp. Referensgruppen är rådgivande och har inte någon direkt styrande funktion. Den akademiska användningen av MONA finansieras i sin helhet av Vetenskapsrådet. Beroende på kostnadsutvecklingen kan finansieringen från Vetenskapsrådet komma att förändras. Vid sidan av den akademiska användningen nyttjas även MONA för andra ändamål. Dessa täcker själva sina kostnader. Nedan ges en översiktlig fördelning av intäkter år 2019, som finansierar kostnaderna för hårdvara, mjukvara och den personal som behövs för att administrera MONA. Kostnader för varje registerhållares arbete för att sammanställa leveranser med mikrodata ingår inte utan debiteras respektive projekt separat.

Översiktlig fördelning av intäkter

Vetenskapsrådet	12 MKR
Övriga	4 MKR
Totalt	16 MKR

3.6 Omfattande men fragmenterat

De korta beskrivningarna ovan visar att det i Sverige pågår ett omfattande arbete med syfte att förse svensk forskning med det e-infrastrukturstöd som krävs för att utveckla och bedriva avancerad forskning. Den direkta finansieringen uppgår idag till drygt 520 miljoner kronor. Till detta kommer de enskilda investeringar i e-infrastruktur som sker på lärosätena. Ytterligare satsningar på e-infrastruktur sker också som en integrerad del av de investeringar som görs i nationell och internationell forskningsinfrastruktur.

Vetenskapsrådet och SUHF är båda övertygade om att investeringarna i e-infrastruktur för forskning måste öka under de kommande åren. Utan en sådan ökning kommer det att bli svårt för Sverige att bibehålla sin position bland de ledande forskningsnationerna.

Ökade investeringar måste dock gå hand i hand med ett effektivare resursutnyttjande och väl övervägda strategiska val. En viktig slutsats från rapporten *An outlook for the national roadmap for e-infrastructure for research* är att det rådande fragmenterade landskapet försvårar upprättandet av en nationell strategi för e-infrastruktur för forskning. Ytterligare en försvårande omständighet är att finansieringen av svensk e-infrastruktur i många delar präglas av kortsiktighet utifrån relativt korta finansieringscykler, och till det kopplad osäkerhet, vilket är olyckligt inom ett område som kräver långsiktig strategi och planering. En konsekvens av detta är att Sverige är illa rustat för att göra strategiskt övervägda investeringar i e-infrastruktur för forskning.

Sverige är på intet sätt unikt i att ha ett fragmenterat e-infrastrukturlandskap. Sverige är inte heller ensamt om att se ett behov av en ökad samordning. Vi har

under arbetets gång studerat och haft kontakt med företrädare för Danmark, Finland, Nederländerna och Norge. Av dessa är Finland det land som kommit längst med att bygga en sammanhållen organisation. Finska IT Center for Science (CSC) startades redan 1971 och har sedan dess utvecklats till en omfattande organisation med ett brett uppdrag att stödja forskning, universitets- och högskolesektorn samt andra delar av den offentliga sektorn. Danmark håller som bäst på att etablera en sammanhållen organisation för forskningens e-infrastruktur. I Norge sker utveckling kring deras datorresurser koordinerade inom UNINETT Sigma2. Nederländerna utreder frågan och brottas med till stor del samma problem som vi identifierat i Sverige. Att frågan om e-infrastruktur är på agendan i många länder är i sin tur en effekt av den simultana utvecklingen av tekniska möjligheter, ökade och allt mer komplexa datamängder, forskningens krav, kraven på öppna forskningsdata och allt mer brännande frågor om den personliga integriteten. Vid en eventuell fortsatt utredning av NYe-Infra finns all anledning att lära från andra länder. Samtidigt måste lösningarna baseras på existerande nationella förutsättningar.

4. Vad behöver lösas?

I takt med att forskningens krav på avancerad e-infrastruktur ökar, växer kraven på långsiktiga och ekonomiskt hållbara lösningar. Utan nya lösningar kommer viss forskning inte att kunna genomföras. I många fall kommer enskilda forskare att själva ta fram egna lösningar mer eller mindre ad-hoc. Risken med detta är att resurser utnyttjas suboptimalt både på kort och lång sikt. Det finns tre avgörande skäl till varför Sverige behöver en sammanhållen organisering av e-infrastruktur för forskning. Dessa är:

- En tydlig hantering och ansvarsfördelning för forskningsdata under hela forskningsdatalivscykeln behöver fastställas på nationell nivå.
- Nät-, lagrings- och beräkningsresurser som möter forskningens behov måste säkerställas.
- Sveriges roll och utbyten av de växande internationella samarbetena måste stärkas.

På ett övergripande plan griper de tre aspekterna in i varandra, därför bör de hanteras i en sammanhållen organisation. Alla tre kräver en nationell strategi. Att Sveriges internationella samarbeten ska utgå från en samlad bild av Sveriges nationella intressen kan tyckas självklart. Högkvalitativ och standardiserad hantering av forskningsdata under hela livscykeln är nödvändigt för att Sverige ska nå målet med öppna forskningsdata år 2026. Investeringar i beräkningsresurser och kapacitet att hantera utveckling av AI kräver en tydlig nationell strategi och ökat internationellt samarbete.

4.1 Hantering av forskningsdata under hela forskningsdatalivscykeln

Sedan 2017 har Vetenskapsrådet i uppdrag att samordna det nationella arbetet med att införa öppen tillgång till forskningsdata. Bakgrunden till uppdraget är bland annat det regeringsuppdrag om att utforma nationella riktlinjer för öppen tillgång till vetenskaplig information som Vetenskapsrådet fick 2013. Målet är att en omställning till öppen tillgång till forskningsresultat inklusive vetenskapliga publikationer, konstnärliga verk och forskningsdata bör vara fullt genomförd senast år 2026. Uppdraget kopplar direkt till den europeiska nivån och det intensiva arbete som nu pågår kring uppbyggandet av European Open Science Cloud (EOSC). För att förverkliga målet krävs processer för att hantera digitala forskningsdata under hela forskningsdatalivscykeln, eller kanske bättre: livscyklarna. Dessa processer inbegriper tre huvudelement:

Främjande av kulturförändring

För att öppen tillgång till forskningsdata i större skala ska bli verklighet krävs utarbetade standarder och principer för hur forskningsdata ska hanteras. För att åstadkomma detta har de så kallade FAIR-principerna utarbetats, principer som stipulerar att forskningsdata och metadata ska vara möjliga att hitta (Findable),

tillgängliga (Accessible), interoperabla (Interoperable), och återanvändningsbara (Reusable). I december 2017 fick Vetenskapsrådet i uppdrag att ta fram kriterier för att kunna bedöma i vilken utsträckning forskningsdata som helt eller delvis tagits fram med offentlig finansiering uppfyller FAIR-principerna. Uppdraget redovisades år 2018.³ Att implementera FAIR-principerna är en omfattande process där förutsättningar, tillämpliga regelverk och behov av stöd skiljer sig mellan olika forskningsområden.

E-infrastruktur för öppen tillgång till forskningsdata

Öppna forskningsdata kräver vidare att det finns teknisk kapacitet att hantera, lagra och tillgängliggöra varierande typer av forskningsdata på ett sätt som motsvarar FAIR-principerna. Olika lösningar krävs exempelvis beroende på typ av forskningsdata (numeriska, visuella, audiella osv.) och inte minst beroende på om forskningsdata innehåller känsliga personuppgifter.

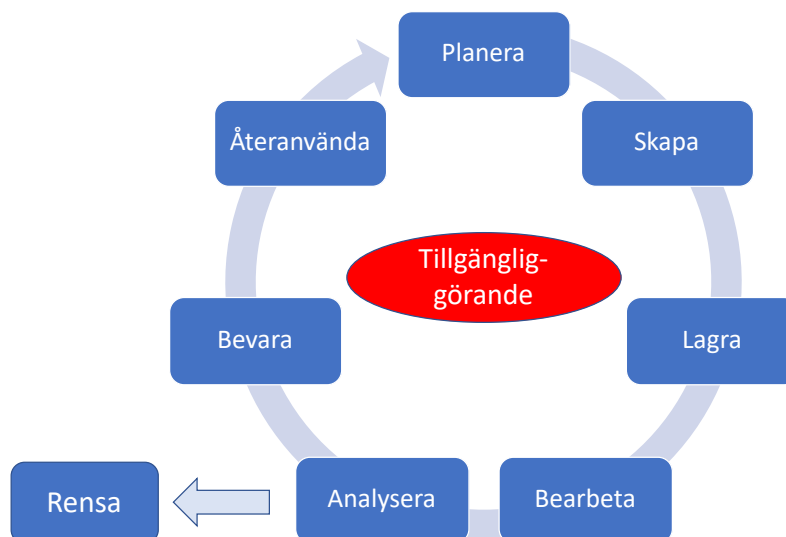
Utveckling av forskningens incitamentsstrukturer

Öppna forskningsdata kan inte åstadkommas utan aktiv medverkan av forskarna. Det är forskare som ska tillämpa principerna och använda de tekniska lösningarna. Det kräver att det finns tydliga riktlinjer för hur FAIR-principer tillämpas och ett förtroende för de tekniska lösningar som erbjuds. Lika viktigt är att principen om öppna forskningsdata accepteras på bred front i forskarsamhället. Två viktiga komponenter för detta är att forskare ser fördelar med att dela med sig av forskningsdata och att de data som delas verkligen kan (åter)användas.

Figur 3 visar en schematisk bild av en typisk livscykel för forskningsdata, en livscykel vars steg är nödvändiga för att data ska kunna tillgängliggöras. Livscykeln börjar med att forskningsprojektet planeras. Vikten av att forskaren redan här beaktar hanteringen av forskningsdata under hela livscykeln kan inte nog understrykas. Centralt är att forskningsdata kan lagras, bearbetas och analyseras på ett effektivt sätt under forskningsprocessen. I samband med att analyserna är klara måste forskaren även fatta beslut om bevarande och eventuell rensning, dvs. beslut om data som inte ska sparas. Skäl till att forskningsdata ska bevaras är att de ska kunna tas fram för att granska forskningsresultat, återanvändas antingen för att följa upp den nyss gjorda forskningen eller för att generera och besvara nya forskningsfrågor med hjälp av existerande forskningsdata. Öppen tillgång till forskningsdata kan även komma det omgivande samhället till gagn, inte minst för innovation och utbildning, men även för att en intresserad allmänhet ska kunna ta del av forskningen i högre utsträckning. En ny organisation för forskningens e-infrastruktur måste utformas så att hela livscykeln stöds. Av det följer att Vetenskapsrådets uppdrag att samordna det nationella arbetet med att införa öppen tillgång till forskningsdata bör överföras till NYe-Infra, då verksamheten utgör en viktig del av samordningen av öppna forskningsdata, på bland annat policynivå, nationellt och internationellt. Detta kommer att utgöra en viktig grund att bygga vidare på för utveckling av incitamentsstrukturer för att göra forskningsdata FAIR och öppna. Nedan kommer de olika stegen att diskuteras något mer detaljerat.

³ Kriterier för FAIR forskningsdata: Redovisning av regeringsuppdraget att ta fram bedömningskriterier för att följa vägen mot ett öppet vetenskapssystem. Rapport 17, 2018

Figur 3. Forskningsdatalivscykeln



I samband med planering av ett forskningsprojekt måste forskaren avgöra vilka datakällor som ska användas. I tablå 1 nedan ges en översikt av de huvudsakliga datakällor som forskare har tillgå.⁴ Vår utgångspunkt är att en ny organisation för forskningens e-infrastruktur under givna förutsättningar ska ges möjlighet att hantera samtliga typer av data.

Tablå 1. Forskningens datakällor

Typ av data	Organisation
Forskargenererade data från experiment och observationer.	Universitet och högskolor
Registerdata	Myndigheter
Vårddata (t.ex. journaluppgifter, kvalitetsregister) och data från klinisk forskning.	Sjukvårdens huvudmän
Övriga data: Kulturarvsdata, icke svenska data, kommersiella data, publika data med mera.	Andra dataägare

⁴ Litteratur och tidigare publicerad forskning är naturligtvis viktiga datakällor. Vi har dock här avstått från att här diskutera öppen tillgång till publikationer.

4.1.1 Forskargenererade data

Forskargenererade data utgör en extremt heterogen kategori där den minsta gemensamma nämnaren är att data har producerats för forskningsändamål. I många fall genereras data vid stora etablerade forskningsinfrastrukturer som exempelvis kärnfysikanläggningen CERN, European Southern Observatory (ESO) och dess avancerade teleskop lokaliserade i Atacamaöknen i Chile eller ICOS, det europeiska systemet för att mäta förekomsten av växthusgaser. Dessa och flera andra forskningsinfrastrukturer som producerar stora datamängder har redan idag av nödvändighet processer för att hantera forskningsdata under hela dess livscykel. I nya stora infrastrukturer som är under konstruktion eller som planeras ses hantering av forskningsdata som en central och integrerad process. Det gäller till exempel neutronkällan ESS som parallellt med att själva anläggningen byggs i Lund bygger upp ett forskningsdatacenter i Köpenhamn. Andra exempel där forskningsdatalivscykeln redan nu hanteras och forskningsdata tillgängliggörs är de europeiska samhällsvetenskapliga undersökningarna European Social Survey och Survey of Health, Ageing and Retirement. Det ska dock understrykas att lösningen för många stora internationella infrastrukturer baseras på att det finns ett internationellt samarbete där nationella resurser samlas. Det kanske mest utvecklade systemet för att samla resurser både vad gäller lagring och beräkning är uppbyggt kring CERN. I framtiden antas EOSC bli viktigt för att samla resurser på EU-nivå för att skapa generella möjligheter till öppen tillgång till e-infrastruktur, tjänster och forskningsdata.

För en betydande del av de forskargenererade forskningsdata är dock situationen mindre klar. I samband med att forskningen planeras är det därför centralt att hela forskningsdatalivscykeln beaktas. *Utan sådan planering kommer målet om öppna forskningsdata inte att kunna nås.* I vissa fall gäller det mycket stora och snabbt ökande datamängder. Här är utmaningen främst att på ett effektivt tekniskt sätt hantera lagring, överföring och analyser av forskningsdata. I andra fall är det inte datamängden i sig som är problematisk utan hanteringen av personuppgifter och andra data som omfattas av sekretess. Det rör till exempel enkätundersökningar, intervjuer, observationer eller experiment där undersökningsobjektet är individer. I vissa fall, framförallt inom medicinsk forskning, är det en kombination av att hantera stora och komplexa databaser som samtidigt innehåller personuppgifter. Till detta kommer forskargenererade data som varken kräver stor lagringskapacitet, resurskrävande beräkningar eller innehåller känsliga uppgifter. Inte desto mindre ska datamängder av detta slag hanteras med hänsyn till hela forskningsdatalivscykeln. Övergripande gäller att förutsättningar ska finnas för att göra forskargenererade data FAIR, oavsett om de lämpar sig för öppen tillgång eller ej.

Vi föreslår att NYe-Infra – en sammanhållen organisation för forskningens e-infrastruktur – ska ges ett nationellt ansvar för att stödja forskare under planeringsfasen av deras forskningsprojekt. Stödet ska därefter gälla hela forskningsdatalivscykeln och täcka datahanteringsplaner, dokumentation/metadata, beständiga filformat samt säker och tillförlitlig lagring.

Ansvaret för forskargenererade data åvilar den organisation, vanligen ett lärosäte, vid vilken forskaren är verksam. Ansvaret för forskningsdata vad gäller säkerhet och arkivering ligger därför också vanligen på ett lärosäte. Vi menar inte att detta ska ändras men att NYe-Infra ska kunna erbjuda stöd, rådgivning och tjänster som

stödjer och underlättar för lärosäten och andra att fullgöra sina åtagande. Myndigheten bör ha i uppgift att följa utvecklingen och ge stöd till forskare i frågor kring forskningsdatahantering, där oklarheter råder i dagsläget, t.ex. licensiering av datamängder och ansvarsförhållanden för forskningsdata under samarbetsprojekt där avtal idag ofta skrivs i fall till fall. Vi ser att det finns stora vinster med en nationell kunskapsuppbyggnad och nationella tekniska lösningar. I den verksamhet som SND nu bygger upp ingår många av de nödvändiga komponenterna. Samtidigt behövs det ökad tydlighet och framförallt långsiktighet i hur forskningsdata ska hanteras och lagras i olika skeden av forskningsprocessen.

Det viktigaste syftet med att upprätthålla FAIR-principerna för forskargenererade data och planera för hela forskningsdatalivscykeln är att forskningsdata ska kunna återanvändas. Att forskningsdata är återanvändningsbara innebär att möjlighet till kontroll av forskningens kvalitet ökar. Vi menar dock att den stora nyttan med att göra forskningsdata återanvändningsbara är att redan insamlade och sammanställda forskningsdata kan användas för att generera och besvara nya forskningsfrågor.

Existerande forskargenererade data kan återanvändas för ny forskning på i princip tre olika sätt.⁵ En existerande databas kan ofta, jämfört med den primära analysen, analyseras på alternativa sätt beroende på exempelvis val av statistiska metoder, hur delar av datamängden kombineras eller helt enkelt för att det finns data som inte fullt ut utnyttjas i den ursprungliga analysen. Ett andra sätt är att kombinera uppgifter från flera likartade databaser. Möjligheten att utnyttja maskininlärning och AI har ökat förmågan att kartlägga systematiska mönster i mycket stora datamängder. Ett tredje sätt är att kombinera data från olika typer av databaser för att på så sätt kunna adressera komplexa vetenskapliga frågeställningar på nya sätt och därmed svara på forskningsfrågor på ett sätt som inte kunnat göras tidigare. Det kan till exempel röra sig om hälsodata som kopplas till mätningar av luftföroreningar kombinerad med vägtrafikdata. För forskningssystemet som helhet är detta potentiellt även en betydande effektivitetsvinst. Vi vill dock understryka att återanvändning av forskningsdata endast i undantagsfall kan ersätta behovet av att generera nya forskningsdata.

I alla tre fallen krävs att forskningsdata är möjliga att hitta, är tillgängliga, interoperabla och återanvändningsbara. Kort sagt, forskningsdata måste så långt det är möjligt vara FAIR. Det betyder i sin tur att varje datamängd måste kompletteras med information om datamängden – metadata och annan dokumentation – enligt fastställda principer. Ett effektivt nyttjande av existerande datamängder kräver att metadata är maskinläsbara och därmed sökbara. I och med det kan enskilda forskare få kännedom om för dem tidigare okända datamängder.

Den verksamhet som idag bedrivs av SND har som huvudsyfte att skapa förutsättningar för att svenska forskningsdata ska leva upp till FAIR-principerna och att tillhandahålla en samlad ingång för att söka efter existerande forskningsdata. Vid sidan av att stödja forskare i processen att göra forskningsdata FAIR ska en NYe-Infra ansvara för en katalog där metadata kopplad till forskargenererade data görs sökbara. Detta är idag en uppgift som i huvudsak ligger på SND men, som framgår nedan, även i Vetenskaprådetets registeruppdrag. Utan en välfungerande katalog- och söktjänst kommer möjligheterna att hitta återanvändbara data vara begränsat.

⁵ Tillgängliggjorda forskningsdata används även inom undervisning och spelar där en viktig roll, inte minst i samband med examensarbeten.

Under forskningsprocessen, från det att forskningsdata skapas till dess de primära analyserna är genomförda, genereras ofta ett antal datamängder som har bearbetats på olika sätt. I experimentella studier är processen ofta tydligt iterativ där analyser leder till att experimentet görs om och nya data genereras. Det är därför långt ifrån givet vilka data som i varje enskilt fall ska bevaras. I fall med mycket stora datamängder och därmed betydande kostnader för lagring, inklusive överföring av data till nya lagringsmedier och filformat, bör också en bedömning av värdet av att spara data göras. Premisserna för vad som ska bevaras varierar mellan olika forskningsområden. Även här behövs nationella standarder och vägledning till forskare. Vi menar att forskarna aktivt bör bidra till att sätta standarder för bevarande och rensning, standarder som måste avpassas till forskningens karaktär och behov samt en ekonomisk rimlighetskalkyl. En ny organisation för forskningens e-infrastruktur ska dock vara drivande för att standarder sätts och att dessa är tillgängliga och kända. En konsekvens av det är att Vetenskapsrådets uppdrag att verka för öppen tillgång till forskningsdata bör överföras till Nye-Infra.

Till sist vill vi med emfas understryka att hantering av forskningsdata under hela forskningsdatalivscykeln kräver forskares aktiva medverkan. Forskare behöver därför ett sammanhållet stöd under hela forskningsprocessen. Tydliga rutiner för dokumentation måste finnas på plats och praktiska lösningar för lagring med mera måste vara tillgängliga. Processen måste även vara meningsfull för den enskilda forskaren. Det kräver att insamling och tillgängliggörande har ett meritvärde och värderas i olika bedömningsprocesser. Om ett arbete läggs ned måste det också kunna visas att forskningsdata faktiskt kan återanvändas på ett meningsfullt sätt. En konsekvens av detta är att somliga discipliner kommer att behöva arbeta med kulturförändringar – att uppvärdera återanvändning av forskningsdata insamlade av andra forskare.

4.1.2 Registerdata

Registerbaserad forskning har funnits länge i Sverige och bedrivs inom exempelvis medicin, folkhälsa och samhällsvetenskap. Genom att koppla samman uppgifter från stora befolkningsgrupper kan forskare bland annat undersöka biologiska och medicinska faktorer, sociala förhållanden och livsföring kopplat till hälsa. Resultat från registerforskning används bland annat för att identifiera riskfaktorer för sjukdomar eller som kunskapsunderlag för politiska beslut. Eftersom forskaren inte behöver samla in uppgifterna själv utan får dem från registren, är forskningen kostnadseffektiv och resultat kan snabbare nås än om forskaren själv ska samla in data. Att svenska myndighetsregister erbjuder unika möjligheter till forskning främst inom medicin och samhällsvetenskap framhålls ofta.

Forskare kan i sin forskning utnyttja ett enstaka register. Det som krävs, vid sidan om att forskaren känner till registret, är en godkänd etikprövning samt att den myndighet som håller registret godkänner att utlämnandet står i överensstämmelse med offentlighets- och sekretesslagen. Ofta kräver forskningen att delmängder av flera olika register kan kombineras. Att detta låter sig göras i register baserade på hela befolkningen är en av anledning till att svensk registerforskning har en världsunik potential. Samtidigt blir processen då mer komplicerad, speciellt om de data som behövs återfinns hos olika myndigheter. Ytterligare en stor potential hos

registerdata är att de är möjliga att kombinera med forskargenererade data som innehåller personuppgifter kopplade till personnummer.

Vetenskapsrådets registeruppdrag syftar dels till att hjälpa forskare att hitta den information som finns i register. För det ändamålet har verktyget Register Utilizer Tool (RUT) utvecklats. RUT hanterar metadata om register: vilka uppgifter finns, för vilka år, vilken population med mera, och gör det möjligt för forskare att söka information medan forskningen planeras. Registeruppdraget syftar också till att stödja forskare och underlätta processen att få tillgång till forskningsdata inklusive etikprövning och prövning om utlämnande. För att möjliggöra koppling mellan forskargenererade data och register planeras RUT även att innehålla information om forskargenererade data i de fall de innehåller personuppgifter, dvs. personnummer som gör det möjligt att koppla dem till registerdata.

En viktig aspekt av användandet av registerdata i forskning är att efter utlämnandet av data är det den mottagande organisationen vid vilken forskaren är verksam som ansvarar för hanteringen och arkiveringen av data. Även i de fall data tillgängliggörs via SCB:s MONA-system så är det lärosätet eller annan organisation där forskaren är verksam som ansvarar för databasen och dess arkivering. Vi menar att detta är problematiskt. En vidare utredning bör beakta möjligheten för NYe-Infra att förvalta andra myndigheters data, inklusive behandling av personuppgifter, i de fall de lämnats ut till forskning. Av det följer att en vidareutveckling av det MONA system som idag finns vid SCB bör byggas upp vid en eventuell NYe-Infra.

4.1.3 Hälsa- och vårddata

Att på ett effektivt sätt tillgängliggöra olika typer av vårddata för forskning är både administrativt och tekniskt utmanande. Samtidigt är forskningspotentialen stor. Redan idag arbetar Vetenskapsrådets registeruppdrag med metadata som tillhör sjukvårdsregionerna, i dagsläget främst kvalitetsregister, och det ställs krav på att de kvalitetsregister som får statligt stöd ska ansluta sig till, dvs. lämna metadata om registret, till registeruppdragets sökverktyg RUT. Även metadata om biobanker och humana prover håller på att integreras i RUT.

I hälso- och sjukvården genereras ökande mängder data inom såväl klinisk som akademisk verksamhet. För utvecklingen av precisionsmedicin är det en förutsättning att dessa data blir forskningsbara, i synnerhet genomikdata (DNA, RNA), proteomikdata och högupplösta avbildningsdata. Majoriteten av sådana data kommer framöver att genereras kliniskt – ca 80 % eller mer enligt prognoser. Möjligheten att kombinera data från sjukvården, såväl klinisk som akademisk, med myndigheters registerdata ger svensk forskning unika möjligheter. Vi föreslår därför att NYe-Infra fortsätter och utvecklar det arbete som nu görs inom Vetenskapsrådets registeruppdrag, genom att sätta standarder för metadata och stödja de olika huvudmännen inom sjukvården i arbetet med att göra data sökbara. Vi vill understryka att ett sådant arbete inte bara är till nytta för forskningen utan även bidrar till att ge de olika huvudmännen bättre överblick över och möjlighet att följa upp existerande databaser. En sådan tillgång ger även möjligheter för utveckling av AI i hälso- och sjukvården.

NYe-Infra bör på sikt även integrera metadata om de forskningsdata som genereras via vetenskapliga kliniska studier, dvs. forskargenererade data inom sjukvården, samt metadata från journalsystem. På samma sätt som för myndigheters

registerdata bör lagstödet utvecklas för att möjliggöra för NYe-Infra att förvalta data och behandla personuppgifter för sjukvårdsregionens data i de fall de lämnats ut till forskning. Ett sådant lagstöd skulle skapa möjlighet för NYe-Infra att skapa säkra beräkningsmiljöer för data innehållande känsliga personuppgifter. Vi återvänder till frågan i kapitel 5 där vi redovisar vårt inriktningsförslag i något mer detaljerat.

4.1.4 Övriga data

Övriga data är med nödvändighet en heterogen kategori. Vi kommer här inte att gå in på övriga data i detalj, bara tydliggöra att vi är fullt medvetna om att forskare använder datakällor som inte täcks av de andra kategorierna i tablå 1, ovan. Det vill säga data som inte är forskargenererade, inte ägs av myndigheter eller sjukvården huvudmän men som ändå har värde för forskningen. Huruvida NYe-Infra är relevant för dessa typer av data måste avgöras från fall till fall. Vi vill dock lyfta att den fortgående digitaliseringen utgör en potential för forskning inom flera områden. Inte minst gäller detta humaniora där digitalisering av bland annat kulturarvsartefakter och litteratur blir allt viktigare och utgör en stor potential för att utveckla forskningen. Det innebär samtidigt att behoven av stöd för analyser av exempelvis komplexa visuella data förväntas öka.

4.2 Tillhandahållande av nät-, lagrings- och beräkningsresurser

4.2.1 Nät

Nätverket är e-infrastrukturens blodomlopp. I takt med att forskningen digitaliseras, datamängder ökar i volym och antal, samt att nationella och internationella gemensamma e-resurser byggs ut ökar kraven på nätförbindelser med hög kapacitet, stor säkerhet och redundans i samband med störningar. Sunet erbjuder idag svenska lärosäten och en rad andra offentliga institutioner ett datornätverk av högsta klass. Vi föreslår att NYe-Infra skall förvalta och utveckla Sunet och svensk medverkan in internationella samarbeten och på så vis möta svensk forsknings behov av överföring av data. En integration av Sunet i NYe-Infra påverkar dock inte de användare av Sunet som är verksamma utanför högskolesektorn. Den verksamheten bör fortsatt upprätthållas.

4.2.2 Lagring

Som framgått ovan är lagring av forskningsdata en central aspekt av modern forskning och utgör en stor utmaning för forskningssystemet. Det är huvudsakligen två grundproblem som måste lösas. För det första finns det ett volymproblem. Forskningen tenderar att generera allt större datamängder. Samtidigt ställs krav på bevarande vilket är en förutsättning för öppna forskningsdata och återanvändning. Vi föreslår att NYe-Infra ska verka för att tillhandahålla resurseffektiv och långsiktig lagring av forskningsdata. Det betyder inte att en ny organisation ska lagra all forskningsdata. Det är ägare av forskningsdata, vanligen ett lärosäte, som ska avgöra hur och var forskningsdata ska lagras. NYe-Infra skall dock kunna erbjuda nationellt koordinerade lagringslösningar när och om sådan efterfrågas av lärosäten eller andra forskningsutövare.

För det andra är det nödvändigt att finna en lösning för lagring av forskningsdata som innehåller personuppgifter eller andra typer av data som är belagda med sekretess. Som framgått ovan kan sådana data vara forskargenererade, härröra från myndigheters register eller sjukvården. I vissa fall kan forskningsdata innehållande denna typ av uppgifter utgöras av en kombination av uppgifter från alla tre huvudkällor. Vi menar att det bör utredas om NYe-Infra kan förvalta och ta fullt ansvar för denna typ av data i de fall de lämnats ut för forskning. Vi föreslår vidare att NYe-infra ska bygga upp kapacitet att lagra och arkivera data utlämnade från myndigheter eller sjukvård för forskningsändamål. Den nya organisationen bör också kunna erbjuda lagring av forskargenererade data som innehåller personuppgifter eller andra sekretessbelagda uppgifter.

Vi föreslår också att NYe-infra ska ge forskare tillgång till att analysera känsliga data via ett utvecklat MONA-system. Det vill säga, forskningsdata som innehåller personuppgifter ska inte distribueras till olika forskningsutövare. I detta sammanhang bör det också utredas om data med personuppgifter som lagras vid NYe-Infra kan ges en särskild legal status som medger absolut sekretess med undantag för forskning.

4.2.3 Beräkning

Allt tyder på att behovet av beräknings- och analysresurser kommer att fortsätta att öka snabbt. Den snabba utvecklingen av artificiell intelligens (AI) och maskininlärning baserad på stora och ofta komplexa datamängder, och av visualiseringstjänster, förstärker dessa behov ytterligare. AI kräver också att beräkningsresurser avsedda för denna verksamhet utvecklas. Vid sidan av de riktigt stora beräkningsbehoven växer efterfrågan på analys- och visualiseringskapacitet. Det kan röra allt ifrån bildserier inom materialvetenskap och komplexa forskningsdata inom livsvetenskaperna, till analys av känsliga individdata. Dessutom ökar efterfrågan från nya användargrupper, utan tidigare vana av att använda beräknings- och analysresurser av detta slag, från bland annat livs- och samhällsvetenskaperna. Detta ställer höga krav på att NYe-Infra tar hänsyn till nuvarande och framtida användares behov i såväl utveckling av resurser och tjänster, som i daglig verksamhet. Vi föreslår att NYe-Infra utifrån en övergripande strategisk planering ska ansvara för att svensk forsknings behov av beräkningsresurser tillgodoses på det mest resurseffektiva sättet, med hänsyn taget till användarnas behov av forskningsnära support. En ny organisation bör också, i enlighet med resonemanget om hantering av data under hela forskningsdatalivscykeln, kunna erbjuda en miljö där beräknings-, analys-, och visualiseringsresurser integreras med föregående och efterföljande faser i forskningsdatalivscykeln på effektivast möjliga sätt. I de fall känsliga data kräver stora beräkningsresurser är detta en nödvändighet.

4.3 Internationella samarbeten

Forskningen blir allt mer internationell och e-infrastrukturen är inget undantag. Sverige medverkar sedan länge i en rad internationella samarbeten inom e-området. Till exempel GÉANT för nätverk och kommunikation, PRACE för beräkningskapacitet, NeIC för samarbete mellan de nordiska länderna. Den internationella nivån har de senaste åren vuxit kraftigt i betydelse. Parallellt tas nu

viktiga initiativ på EU-nivå, framförallt gäller det European Science Cloud (EOSC) och The European High-Performance Computing Joint Undertaking (EuroHPC).

EOSC syftar till två saker. För det första att verka för öppna forskningsdata vilket sammanfaller med Vetenskapsrådets uppdrag att verka för att svenska forskningsdata är öppet tillgängliga år 2026. Sverige är därför engagerat i EOSC-arbetet, där verksamheter som Svensk nationell datatjänst (SND), SNIC och Vetenskapsrådets registeruppdrag kan ses som nationella resurser till EOSC. Det andra syftet med EOSC är att samordna och tillgängliggöra digitala resurser och tjänster för forskning, framförallt beräkningsresurser och lagringskapacitet. Vad gäller beräkningsresurser finns en direkt koppling till EuroHPC-satsningen. Vetenskapsrådet finansierar sedan i år (2019) det svenska deltagandet i ett EuroHPC-samarbete lett av Finland. Inom det samarbetet kommer LUMI, en så kallad pre-exascaledator, att byggas i finska Kajaani. Satsningen är unik då det är första gången flera länder, åtta stycken, gemensamt bygger en samlokaliserad dataresurs med syfte att förse länderna med beräkningskapacitet som de enskilt inte har förmåga att bygga upp.⁶ Detta ger möjlighet för Sverige att utveckla en strategi för kompletterande beräkningsresurser och anpassade lösningar utifrån svenska forskares behov.

Både EOSC- och EuroHPC-satsningarna visar att digitala e-infrastrukturer inte längre är nationella angelägenheter utan står i centrum för satsningar på EU-nivå. Samtidigt kan det internationella samarbetet endast drivas om det finns en fungerande nationell verksamhet. EuroHPC är ett komplement till, inte en ersättning för, svensk beräkningskapacitet och EOSC i sin tur är inte möjligt utan nationella funktioner. För att Sverige fullt ut ska kunna bidra till och, ännu viktigare, dra nytta av de initiativ som har tagits och tas på den europeiska nivån krävs en samlad svensk strategi och kopplad till den en handlingsplan. Vi menar att en ny organisation för forskningens e-infrastruktur på ett samlat sätt ska kunna företräda Sverige i internationella samarbeten inom forskningsdata- och e-infrastrukturuområdet. Möjligheterna för Sverige att påverka internationellt och att dra nytta av samarbetet skulle på så sätt öka.

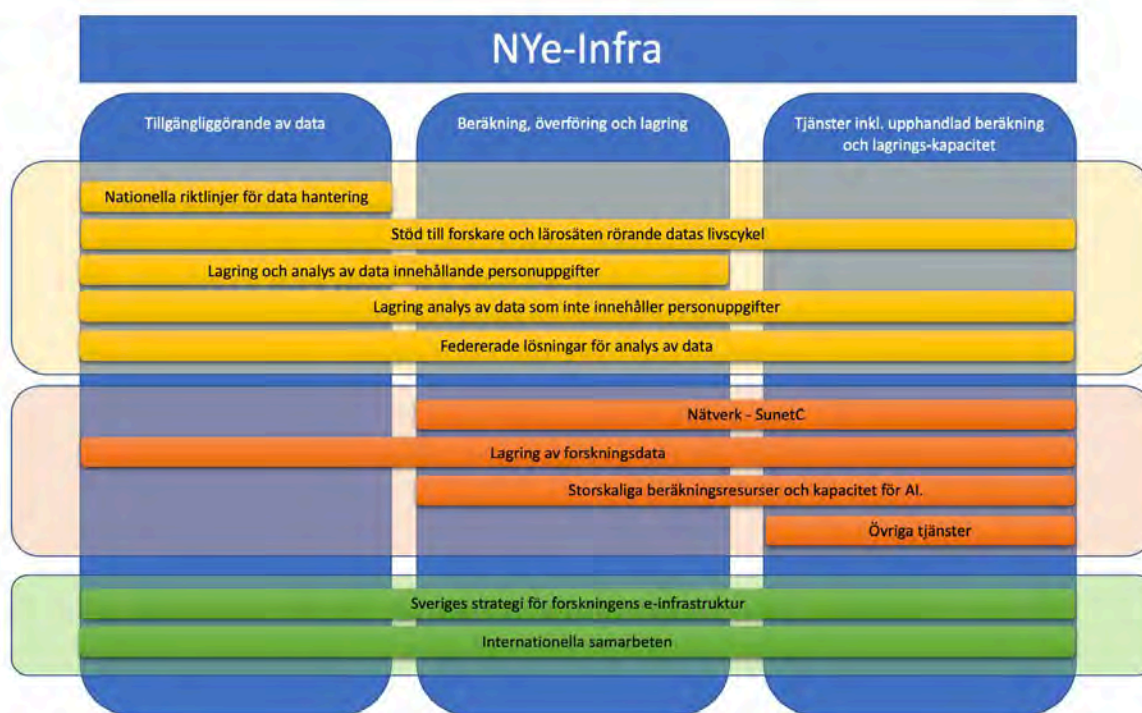
⁶ Samarbetet omfattar för närvarande Finland, Sverige, Danmark, Norge Schweiz, Tjeckien, Belgien och Nederländerna.

5. En något mer detaljerad beskrivning av inriktningsförslaget

I kapitel 2 presenterade vi övergripande vårt förslag om en ny svensk e-infrastruktur för forskning. Avsikten med detta kapitel är att diskutera förslaget något mer detaljerat kopplat till den tänkta organisationsstrukturen. Åter vill vi understryka att det vi framför är ett inriktningsförslag som kräver vidare utredning. Vi är också medvetna om att det med säkerhet finns aspekter som vi förbisett.

I figur 4 har den initiala idén med tre NYe-Infra-avdelningar (se figur 1) kompletterats med en likaledes tentativ matrisorganisation. Vi har delat in NYe-Infras uppgifter i tre kategorier – hantering av data, nät- lagrings- och beräkningsresurser samt övriga tjänster. Inom var och en av dessa kategorier har vi specificerat uppgifter som diskuteras i den ordning de står uppställda i figur 4. Som framgår av figur 4 menar vi att de flesta av dessa uppgifter berör mer än bara en av NYe-Infras tänkta avdelningar vilket i sig är den grundläggande orsaken till varför vi anser att en sammanhållen organisation är nödvändig. Nedan redogör vi för vad vi i detta skede ser som de viktigaste komponenterna i matrisorganisationen.

Figur 4. Tentativ matrisorganisation för NYe-Infra



5.1.1 Hantering av data

NYe-Infra bör få ansvar att verka för nationella riktlinjer för datahantering och rådgivning vad gäller legala och praktiska förutsättningar för att hantera data under

dess livscykel. Standarder för metadata och lagring måste så långt som är möjligt koordineras på nationell nivå och synkroniseras med andra intressenter nationellt och internationellt. NYe-Infra bör som organisation ges ansvar att utfärda beständiga identifikatorer (PID – Persistent Identifiers), exempelvis DOI:er (Digital Object Identifiers), vilket är ett av kriterierna för att data ska vara FAIR. Forskare måste kunna få stöd i frågor som rör ägande- och ansvarsförhållanden vid samarbeten mellan både nationella parter och internationella parter. Detta kan skilja sig åt under olika faser av forskningsdatalivscykeln, och inbegriper ekonomiskt ansvar och forskarens individuella rättigheter och skyldigheter för tillgängliggörande av forskningsdata. En aktuell fråga gäller licensiering av datamängder, där det finns risk att en forskare applicerar licenser, till exempel Creative Commons Non-Commercial (CC-NC) som begränsar användning av material för kommersiella ändamål men som står i motsats till såväl juridiska ramverk som till lärosätens, forskningsinfrastrukturers, och forskningsfinansiärers strategier för öppen tillgång till forskningsdata. Synkronisering med andra myndigheter med beröringspunkter, exempelvis Etikprövningsmyndigheten, DIGG och Riksarkivet är också centralt.

NYe-Infra bör ta ett nationellt ansvar för stöd till forskare och lärosäten under hela forskningsdatalivscykeln

Detta rör, under planeringsskedet (se figur 3), utformande av datahanteringsplaner inkluderande standarder för metadata, lagring och omrensning. Mycket av detta utgör det basala långsiktiga arbetet för tillgängliggörande av forskningsdata. I de fall forskare är i behov av lagrings- och beräkningsresurser bör också detta ingå i planeringen och NYe-Infra bör kunna ge råd och stöd. Det samma gäller övriga tjänster. NYe-Infra kommer på så sätt att kunna fungera som en “one stop shop” för hela planeringsprocessen. Observera dock att ansvaret för forskningsdata och forskningens utförande inte primärt åligger NYe-Infra. Det är alltid det lärosäte eller organisation som forskaren är verksam vid som har ansvar för data under hela forskningsdatalivscykeln. Vi menar inte heller att forskningsdata primärt skall lagras hos NYe-Infra. Även fortsättningsvis kommer forskningsdata i huvudsak att lagras och arkiveras hos främst lärosäten. NYe-Infra kommer dock att verka för att de forskningsdata som finns vid lärosätena kommer att vara sökbara och återanvändningsbara via fortsatt utveckling av de tjänster som idag hanteras av SND och Registeruppdraget. I den mån lärosäten och forskare efterfrågar lagrings- och beräkningstjänster skall dock NYe-Infra kunna erbjuda dessa.

Lika lite som att NYe-Infra kommer att kunna eller ens bör ha ansvar för all lagring och beräkning av alla forskningsdata kan NYe-Infra ge service åt varje enskild forskare. Att det övergripande stödet också vänder sig till lärosäten och andra forskningsutövare är därför centralt då stödet till enskilda forskare måste ske lokalt. Vi ser därför framför oss att det arbete som SND inlett med lokala forskningsdatastödfunktioner kommer att fortsätta och vidareutvecklas. Vi tror också att NYe-Infra kommer att ha bättre förutsättningar för detta genom att bidra med kompetens och tjänster kopplade till alla tre avdelningarna och rörande alla delar av datas livscykel.

NYe-Infra ska vara den centrala aktören för tillgängliggörande av forskningsdata

Idag har både SND och Vetenskapsrådets registeruppdrag metadata och verktyg som ger forskare möjlighet att under planeringsfasen söka efter både forskaregenererade data, registerdata och sjukvårdsdata. Den främsta metoden att öka mängden forskaregenererade data som är sökbara är det stöd till forskare under forskningsdatalivscykelns som diskuteras ovan. När det gäller registerdata och sjukvårdens data krävs att det arbete som utförs inom ramen för registeruppdraget inte bara fortgår utan också utökas. Målsättningen är att NYe-Infra ska fungera som en "one stop shop" för forskare som söker efter existerande data. Som nämnts tidigare kommer data i huvudsak inte att lagras hos NYe-Infra. Data kommer även fortsättningsvis att vara distribuerade bland lärosäten, myndigheter och sjukvården. Att NYe-Infra kan erbjuda lagringstjänster till lärosäten förändrar inte detta, legalt är forskningsdata fortfarande lärosätets ansvar. NYe-Infra kommer endast att ge tillgång till metadata och verktyg som gör dessa åtkomliga och sökbara. Som nämnts ovan betyder det att forskare som vill nyttja data, när så är nödvändigt, behöver ansöka om etiskt tillstånd för forskningens utförande samt utlämning prövad enligt offentlighets- och sekretesslagen av den utlämnande organisationen. NYe-Infra bör dock kunna ge stöd under utlämnandeprocessen. Idag finns sådant stöd att tillgå både via Vetenskapsrådets registeruppdrag och SND. I de fall ett utlämnande av data kräver resurser i form av lagring och beräkning bör NYe-Infra kunna hantera detta under tillgängliggörande-processen, vilket gör att även tillgängliggörande spänner över NYe-Infra's alla tre avdelningar. Vi menar också att det bör övervägas att NYe-Infra ska kunna ge juridisk rådgivning. Det kan exempelvis gälla då forskare vill ha tillgång till data med känsliga personuppgifter, data som är skyddade av upphovsrätt eller ägs av kommersiella aktörer.

Vårt förslag innebär alltså att NYe-Infra som regel inte skall ska vara förvaltare av data. *Vi anser dock att NYe-Infra bör ges ett särskilt ansvar för forskningsdata innehållande känsliga personuppgifter från myndighetsregister och sjukvårdsdata.* Flera statliga utredningar har undersökt de legala förutsättningarna för att bättre utnyttja myndigheternas register och sjukvårdens data för forskning.⁷ En utredning av möjligheten att inrätta en ny organisation för svensk e-infrastruktur bör ta hänsyn till förutsättningarna för den nya organisationen att förvalta och hantera personuppgifter för forskningsdatabaser från register och sjukvårdsdata. Vi menar att NYe-Infra kan ge de organisatoriska förutsättningarna för att ge legalt stöd till de förslag om forskningsdatabaser som tidigare framförts i statliga utredningar. Vår grundtanke är i grova drag följande:

- När ett beslut om utlämnande av data från register eller sjukvårdsdata har fattats, efter sedvanlig etikprövning och prövning enligt offentlighets- och sekretesslagen, överförs den definierade datamängden till NYe-Infra.
- I de fall data härrör från en enda huvudman utgör de överförda data forskningsdatabasen. Har forskaren tillstånd att komplettera med ytterligare persondata ingår även personnummer i de överförda data.

⁷ Rätt att forska – Långsiktig reglering av forskningsdatabaser (SOU 2018:36), Unik kunskap genom registerforskning (SOU 2014:45)

- I de fall data härrör från flera huvudmän överförs data från var och en till NYe-Infra inklusive personnummer. NYe-Infra sammanställer sedan forskningsdatabasen.
- I de fall personnummer förekommer, skapas en unik identifikationsnyckel för varje forskningsdatabas. Personnumren kan därefter skiljas från forskningsdatabasen och tillsammans med nyckeln förvaras av NYe-Infra separat från forskningsdatabasen.
- Ansvaret för data som på detta sätt överförs till NYe-Infra övergår i sin helhet till NYe-Infra. NYe-Infra kommer alltså att förvalta, vara ansvarig för personuppgifter och ansvara för arkivering av data. Ansvaret för grunddata ligger helt och hållet kvar på den ursprungliga huvudmannen.
- Data som förvaltas av NYe-Infra stannar hos NYe-Infra, någon fysisk distribuering av forskningsdatabaser från NYe-Infra till lärosäten och andra forskningsutövare kommer inte att ske.
- Forskares tillgång till data hanteras via fjärranslutning. Det betyder att NYe-Infra tar över och vidareutvecklar dagens MONA-system.

Detta upplägg löser en rad praktiska problem och vår förhoppning är att det också kan bidra till tydlighet vad gäller legala förutsättningar. Det senare måste dock noggrant utredas. Praktiskt ser vi främst följande fördelar:

- Stärkt säkerhet och skydd av personlig integritet i de fall känsliga persondata används för forskning. Data kommer inte att fysiskt distribueras till flera olika huvudmän vilket minskar risken för dataläckage. Det betyder också att forskare, lärosäten och andra forskningsutövare inte behöver bygga upp separata system för att hantera och arkivera data med känsliga personuppgifter samt att ansvarsfördelningen för datasäkerhet tydliggörs.
- NYe-Infra kommer att kunna implementera de rutiner som behövs för att uppfylla FAIR-principerna för samtliga forskningsdatabaser som förvaltas av NYe-Infra. Detta kommer att skapa förutsättningar för att redan existerande forskningsdatabaser kan återanvändas. Praktiskt kommer det att vara möjligt att kombinera forskningsdatabaser som förvaltas av NYe-Infra. Den stora vinsten med detta är att myndigheter och sjukvården i mindre grad belastas med ärenden om utlämnande av data. Det betyder i sin tur att mängden utlämnade databaser borde kunna minskas vilket i sig är en fördel, inte minst utifrån ett datasäkerhetsperspektiv. Vi menar också att det finns effektivitetsvinster även från forskarnas perspektiv då det är mycket tidseffektivt att kunna använda en redan existerande och väldokumenterad databas jämfört med den omfattande processen att sätta samman en helt ny forskningsdatabas. De legala förutsättningarna för återanvändning av forskningsdatabaser måste dock utredas.
- Myndigheters register och sjukvårdens data är i viss utsträckning föränderliga i samband med att det sker uppdateringar och rättelser. Det betyder att svårigheter kan uppstå i fall då forskning skall reproduceras eller kontrolleras. De svårigheterna undviks i och med att de databaser som förvaltas av NYe-Infra inte kommer att förändras.

NYe-Infra bör också erbjuda forskningsutövare att lagra forskargenererade data med känsliga personuppgifter. Även i dessa fall bör access ges via fjärranslutning till de forskare som har vederbörligt tillstånd att analysera data. Redan idag erbjuder SNIC sådana möjligheter via det så kallade SNIC-SENS som hanterar känsliga genomdata från National Genomics Infrastructure (NGI).

Vi menar att det är en fördel att kunna hantera forskningsdatabaser i en sammanhållen nationell organisation med tydlig praxis och hanteringsordning. Inte minst viktigt är det att dokumentation, och uppdatering av forskningsdatabaserna sker i takt med att data bearbetas, vilket i sin tur är en förutsättning för att sammanställa sökbara metadata av god kvalitet. Dessutom kan lagring och analys ske i en sammanhållen och säker miljö. För att hantera känsliga data måste med största säkerhet NYe-Infra skapa internt uppbyggda tjänster, då kommersiellt tillgängliga beräknings- och lagringsresurser torde vara mindre lämpliga utifrån både ett förtroendeperspektiv och ett legalt perspektiv.

Merparten av forskargenererade data innehåller inte känsliga personuppgifter

Även för dessa gäller att de ska hanteras under hela forskningsdatalivscykeln, såsom beskrivits ovan. Även för denna typ av data kommer NYe-Infra att spela en viktig roll. Anta att en forskare får tid att genomföra ett experiment vid ljussynkrotronanläggningen Max IV. Experimentet genererar stora mängder data och enligt gällande praxis kommer forskaren att kunna bearbeta och lagra data vid Max IV i tre månader efter det att experimentet slutförts. Därefter är det forskarens och i förlängningen forskarens lärosätes ansvar att hantera data som genererats, och att bereda möjligheter för vidare analyser. Vi menar att NYe-Infra bör kunna erbjuda nationellt koordinerade tjänster rörande lagring och beräkning som gör det möjligt för forskare att planera för hela forskningsdatalivscykeln. På en nationell nivå torde det leda till ett effektivare resursutnyttjande. För denna typ av forskningsdata bör också NYe-Infra ha möjlighet att utnyttja kommersiellt tillgängliga resurser.

Resonemangen ovan har utgått från att forskningsdata utgörs av en definierad datamängd som lagras på en fysisk plats. Ett alternativ till detta är federerade lösningar. Mycket förenklat innebär en federerad lösning att forskaren ges möjlighet att analysera flera likartade databaser utan att dessa sammanförs till en fysiskt sammanhållen databas. Anta att det i varje sjukvårdsregion finns en databas rörande behandling av en specifik sjukdom. I en traditionell lösning skulle forskaren begära ut data från varje region och konstruera en ny databas. I en federerad lösning kommer data ligga kvar hos de enskilda huvudmännen, access ges endast i analysögonblicket och forskaren kan endast ta del av analysresultaten, aldrig själva databasen. Utvecklingen av federerade lösningar har pågått under en följd av år och förväntas spela en allt större roll i framtiden. Vi vill dock understryka att federerade lösningar är beroende av att de enskilda databaserna uppfyller FAIR-principerna och kraven på interoperabilitet. Federerade lösningar erbjuder en alternativ lösning när det gäller tillgång till data men samtidigt ställer de krav på att data hanteras likformigt vad gäller standarder för dokumentation, metadata och lagring. Vi menar att federerade data i närtid framförallt kan spela en betydande roll i internationella forskningssamarbeten där data med känsliga personuppgifter från flera länder

5.1.2 Nät-, lagrings- och beräkningsresurser

Kapaciteter vad gäller nät-, lagrings- och beräkningsresurser kan inte separeras från frågor som rör hantering av forskningsdata. Samtidigt vill vi markera att tillgodoseende av forskningens behov av dessa resurser i sig är en komplex verksamhet som kräver både långsiktig strategisk planering och långsiktiga finansiella åtaganden.

Det svenska universitetsnätverket uppdaterades under 2018 till vad som kallas SunetC vilket innebär att svenska lärosäten och övriga användare har tillgång till internet av högsta klass, stor driftssäkerhet och betydande redundans vid eventuella driftstörningar. En viktig uppgift för NYe-Infra är att fortsätta driften av SunetC, utveckla detta och planera för dess uppdatering.

Kraven på att kunna lagra forskningsdata ökar snabbt. Det beror dels på att allt större datamängder genereras, dels på kravet på öppen tillgång, dels på det därav följande kravet att hantera datamängder så att de uppfyller FAIR-principerna och på så vis underlättar öppen tillgång till forskningsdata. Som anförts ovan bör NYe-Infra få möjlighet att lagra och förvalta data med persondata som helt eller delvis består av uppgifter från myndigheters register eller sjukvårdsdata. För detta måste NYe-Infra tillhandahålla intern lagringskapacitet. I andra fall kan NYe-Infra tillhandahålla lagringskapacitet åt lärosäten och andra forskningsutförare men då i form av en tjänst. NYe-Infra kommer då inte att förvalta data utan enbart upplåta lagringsutrymme, ansvaret för förvaltning av data åvilar den som utnyttjar tjänsten. Beroende på typ av data, vad som efterfrågas samt naturligtvis vad som är mest kostnadseffektivt kan lagring erbjudas på interna resurser eller på kommersiellt upphandlade resurser. Kravet på lagringsresurser kommer att vara mycket varierat både med hänsyn till volym och typ av data. Vi föreställer oss också att behoven varierar beroende på om och i så fall hur datalagringen är kopplad till analysbehov.

Att tillhandahålla storskaliga beräkningsresurser kommer att vara en av de viktigaste uppgifterna för NYe-Infra. Idag är de svenska nationella beräkningsresurserna distribuerade på sex olika HPC-centra och verksamheten hålls samman av SNIC. Sverige är också via SNIC engagerade i flera internationella samarbeten, t.ex. PRACE, WLCG, EuroHPC och NeIC. Forskningens krav på storskalig beräkning ökar snabbt och i tandem med den tekniska utvecklingen av HPC-resurser. Den effektiva drifttiden för avancerade HPC-resurser beräknas idag vara 3 till 5 år. Efter det behöver de ersättas. En viktig uppgift för NYe-Infra blir därför att strategiskt bedöma svensk forsknings behov av HPC-resurser och långsiktigt planera för hur dessa ska mötas. Inte minst gäller det krav på resurser kopplade till utvecklingen av AI och maskininlärning. Centrala överväganden gällande exempelvis vilka resurser Sverige skall införskaffa nationellt och vilka som byggs i samarbete med andra länder. För de resurser som byggs nationellt krävs noggranna överväganden om lokalisering då det måste avgöras om Sverige ska fortsätta ha flera HPC-centra eller om vi ska följa Finlands exempel och koncentrera resurserna till en lokalitet.

5.1.3 Övriga tjänster

Som framgått ovan erbjuder framförallt Sunet redan idag olika typer av stöd i form av tjänster. Vi ser att denna typ av tjänster även fortsättningsvis kommer att vara av vikt.

5.1.4 Nationell strategi och internationella samarbeten

En viktig slutsats från rapporten *An outlook for the national roadmap for e-infrastructure for research* är att det fragmenterade landskapet försvårar upprättandet av en nationell strategi för e-infrastruktur för forskning. En sådan strategi bör omfatta en tydlig riktning för beräkningsresurser, lagring av olika typer av data (vilket betyder olika lagringslösningar), överföring och kommunikation, hantering av forskningsdata och strategi för öppna forskningsdata. Vi föreslår att NYe-Infra ska få ansvar för att utarbeta en sådan strategi.

Behovet av en svensk strategi ökar i takt med att den internationella arenan blir viktigare. Sverige måste helt enkelt förhålla sig till utvecklingen av europeiska initiativ som exempelvis EOSC och EuroHPC. I de internationella sammanhangen är det av central betydelse att Sverige är tydligt gällande vem som företräder och talar för Sverige samt att den som för talan tar sin utgångspunkt i en svensk strategi för e-infrastruktur för forskning. Vi menar att NYe-Infra bör ges i uppdrag och få mandat att företräda Sverige i frågor som rör internationellt e-infrastruktursamarbete.

5.1.5 Organisationsform och ledningsstruktur

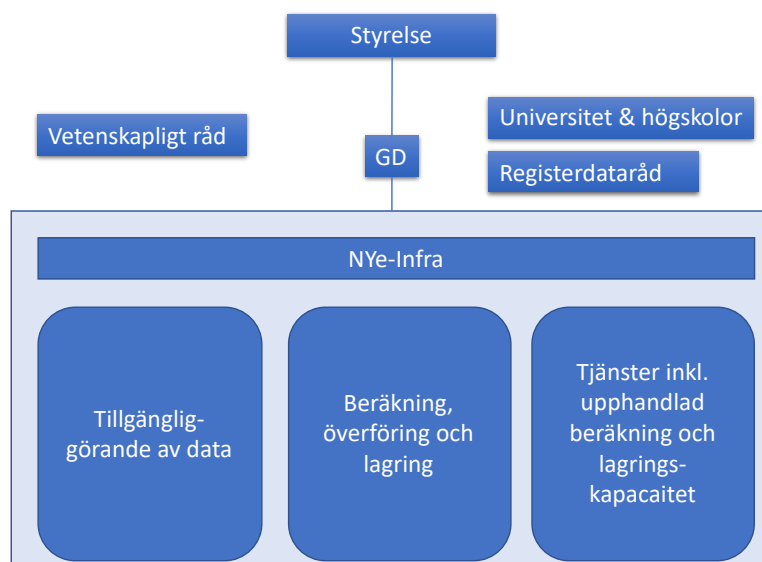
Regeringen bör tillsätta en utredning med syfte att åstadkomma en ny sammanhållen organisation för forskningens e-infrastruktur. En fråga för utredningen är att utröna vilken typ av organisationsform som är mest lämpad för detta ändamål. Bland de närliggande länder som vi tittat på är det finska CSC som är det tydligaste exemplet på en sammanhållen organisation. CSC är organiserat som ett icke vinstdrivande aktiebolag med särskilt uppdrag. Aktieägare i bolaget är finska staten och finska universitet och högskolor. För den organisation vi skisserat ovan är denna organisationsform problematiskt. Den primära orsaken är att en central uppgift för NYe-Infra är att hantera och förvalta data från myndigheter och sjukvård som innehåller personuppgifter, när dessa data används för forskning. Vi tror att NYe-Infra måste vara en myndighet för att det ska vara möjligt. Vi kan också konstatera att CSC inte kan hantera data på det sätt som vi ovan diskuterat och Finland inrättar från och med 2020 en ny myndighet, Myndigheten för digitalisering och befolkningsdata, för detta ändamål.

I det fall beslut tas om att skapa en sammanhållen organisation för forskningens e-infrastruktur kommer denna att ha en central roll för svensk forskning. Att på bästa sätt tillgodose svensk forsknings behov kommer att vara den främsta uppgiften för ett eventuellt NYe-Infra, vilket ställer stora krav på ledning och styrning. Vi presenterar inte en detaljerad bild av hur vi ser på den ledningsstruktur vi anser att ett framtida NYe-Infra bör ha, utan pekar enbart på några komponenter som vi menar bör ingå i ledningsorganisationen.

Vi anser att NYe-Infra bör vara en styrelsemyndighet. Skälet till det är att NYe-Infra kommer att ha central roll i det svenska forskningssystemet med ansvar för att formulera och genomföra strategiska beslut som kännetecknas av stor komplexitet och påverkar hela universitets- och högskolesektorn samt en rad andra aktörer. Vidare föreslår vi att NYe-Infra ska företräda Sverige i internationella samarbeten rörenade e-infrastruktur för forskning. För att fullgöra dessa uppgifter behöver NYe-Infra en styrelse med god förståelse av forskningens behov av e-infrastruktur. Den

operativa ledningen av NYe-Infra bör, givet myndighetsformen, åvila en generaldirektör (GD).

Figur 5. Indikativ skiss av ledningsstruktur



Till NYe-Infra bör också kopplas ett vetenskapligt råd för att stödja styrelsen i arbetet med att ta fram en långsiktig strategi för forskningens behov av e-infrastruktur. Till det vetenskapliga rådet bör vetenskapligt rådgivande grupper bestående av användare av NYe-Infra knytas. Vetenskapsrådets registeruppdrag har idag ett registerdataråd knutet till sig, sammansatt av representanter för de stora registerhallarna, Statistiska centralbyrån och Socialstyrelsen, Datainspektionen, biobankerna samt forskarrepresentanter. Detta råd bör behållas, samtidigt som representation från sjukvårdens huvudmän bör inkluderas i rådet. Vi anser att registerdatarådet har en viktig roll att spela i relation till NYe-Infra och tillgängliggörande av register- och sjukvårdsdata för forskning. Vi menar vidare att svenska lärosäten måste ha en tydlig arena för att kommunicera med NYe-Infra.

De komponenter som återges i figur 5 ovan, återspeglar relativt väl de delar som återfinns i finska CSC:s ledningsstruktur. CSC har också utvecklat ett tydligt årsvis schema för hur de olika intressenterna ska interagera och vilka roller de har. Mycket erfarenhet kan hämtas från det finska exemplet, i det fall regeringen väljer att utreda en ny organisation för forskningens e-infrastruktur.

5.1.6 Finansiering

Som visades i avsnittet om den nuvarande organiseringen av svensk e-infrastruktur är finansieringen av befintliga verksamheter heterogen. Vi menar dock att finansiering av en ny organisation för forskningens e-infrastruktur bör följa några enkla principer. I de fall NYe-Infra realiseras bör Vetenskapsrådets nuvarande finansiering av SNIC, Sunet, SND, registeruppdraget och MONA överföras till NYe-Infra. Samtidigt kvarstår de äskanden om ytterligare medel till e-infrastruktur som Vetenskapsrådet gör i sitt inspel till regeringen inför 2020 års forskningsproposition. Dessa medel bör också tillföras NYe-Infra.

Även i relation till lärosätenas kostnader bör den initiala finansieringen av NYe-Infra vara kostnadsneutral. De kostnader som lärosäten har för Sunet, nätverk och tjänster, bör kunna hanteras enligt nuvarande princip. Det innebär att också att övriga användare av Sunet inte primärt påverkas. De kostnader som lärosäten har för SNIC och SND bör också tillföras NYe-Infra. Hur det långsiktigt ska göras bör dock utredas mer i detalj. NYe-Infra bör ges möjlighet att vara mottagare av bidrag från forskningsfinansiärer och andra finansiärer inklusive EU. Då vi förväntar oss att NYe-Infra kommer att ta ett ökat ansvar för lagring av data, både data som NYe-Infra förvaltar och data som NYe-Infra lagrar som tjänst bör principer för avgiftsbeläggning klarläggas i en eventuell utredning.

6. Lista över förkortningar och termer

Term	Beskrivning
CC-NC	Creative Commons Non-Commercial
CSC	IT Center for Science, Finland
DAU	Data Access Unit
DIGG	Myndigheten för digital förvaltning
DOI	Digital Object Identifier
e-infrastruktur för forskning	Med e-infrastruktur för forskning avses resurser för lagring, överföring, tillgängliggörande, beräkning och analys av digitala data. Inom begreppet ryms också de kompetenser och organisationer som är nödvändig för att forskningen ska kunna tillgodogöra sig befintliga resurser och existerande data.
EOSC	European Open Science Cloud
ESO	European Southern Observatory
ESS	European Spallation Source
EuroHPC	The European High-Performance Computing Joint Undertaking
FAIR(-principerna)	En mängd principer som stöttar att datamängder och relaterade metadata blir Findable, Accessible, Interoperable, Reusable
GÉANT	Gigabit European Advanced Network Technology. Ett europeiskt stamnät med hög bandbredd för forskning och utbildning.
HPC	High-Performance Computing
ICOS	Integrated Carbon Observation System
KAW	Knut och Alice Wallenbergs Stiftelse
Ladok	Ett studentadministrationssystem, utläses Lokalt ADB-baserat dokumentationssystem

Term	Beskrivning
Max IV	En svensk forskningsinfrastruktur för synkrotronljusforskning
MONA	Microdata Online Access
MSB	Myndigheten för samhällsskydd och beredskap
NeIC	Nordic eInfrastructure Collaboration
NGI	National Genomics Infrastructure
NORDUnet	Samarbete mellan nationella forsknings- och undervisningsnätverk i de nordiska länderna
NyA	System för antagning till högskoleutbildning i Sverige
NyE-Infra	En ny sammanhållen organisation för forskningens e-infrastruktur
PID	Persistent Identifiers
PRACE	Partnership for Advanced Computing in Europe
Prisma	Fortes, Formas och Vetenskapsrådets gemensamma ansöknings- och ärendehanteringssystem
PTS	Post- och telestyrelsen
Registeruppdraget (Vetenskapsrådets registeruppdrag)	Vetenskapsrådets uppdrag från regeringen för att öka tillgängligheten till, och därmed möjligheten att beforska registerdata
RISE	Research Institutes of Sweden
RUT	Register Utilizer Tool
SCB	Statistiska centralbyrån
SND	Svensk nationell datatjänst
SNIC	Swedish National Infrastructure for Computing
SNIC-SENS	Ett SNIC-projekt med mål att tillhandahålla tjänster för säker hantering av känsliga data

Term	Beskrivning
SUHF	Sveriges universitets- och högskoleförbund
SUNET	Svenska universitetsnätverket
SUNETC	SUNETs nätverk
SWAMID	Identitetsfederation för de flesta lärosäten, forskningsinstitut och andra myndigheter inom Sveriges forsknings- och utbildningssektor
TLV	Tandvårds- och läkemedelsförmånsverket
UNINETT Sigma2	Norges nationella infrastruktur för HPC och datalagring
URFI	Universitetens referensgrupp för forskningsinfrastruktur
WLCG	Worldwide LHC (Large Hadron Collider) Computing Grid

Det är angeläget att skapa förutsättningar för Sverige att utveckla en e-infrastruktur för forskning som möter framtidens behov, tar till vara på de möjligheter Sverige har som forskningsnation och medverkar till att göra forskningsdata öppet tillgänglig.

För att åstadkomma detta föreslår Vetenskapsrådet tillsammans med Sveriges universitets- och högskoleförbund en ny organisation för forskningens e-infrastruktur.

Vetenskapsrådet
Västra Järnvägsgatan 3
Box 1035, 101 38 Stockholm
Tel 08-546 44 000
vetenskapsradet@vr.se
Vetenskapsrådet.se