

Projektplan: Forskares behov av lagringslösningar för forskningsdata – en pilotstudie av SNIC/SUNET/SND.

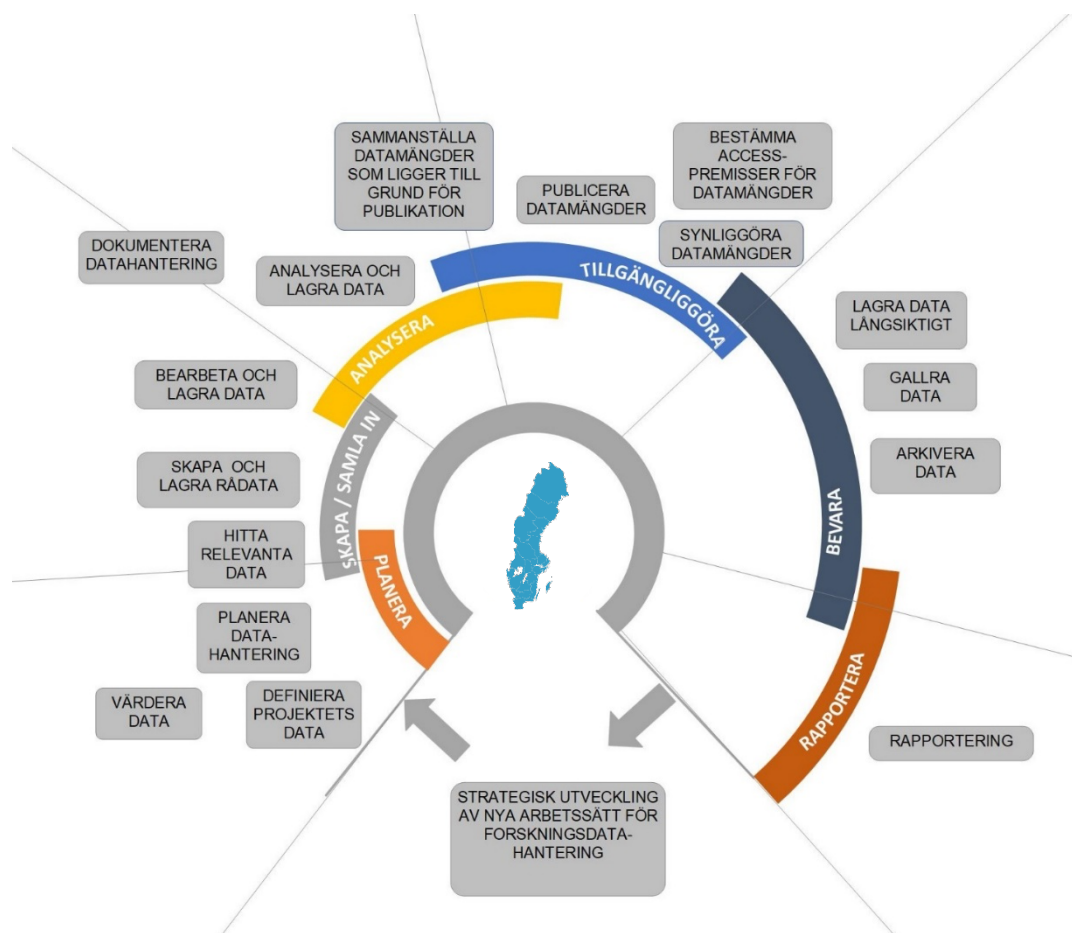
Styrgruppsmöte 2021-09-17

Bakgrund

Det svenska e-infrastrukturlandskapet är fragmenterat och, trots rekommendationer från beslutsfattare så ökar fragmenteringen. De nationella aktörer som har uppdrag och/eller mandat som rör forskningsdata är främst SNIC, SND, och Sunet. Dessa aktörer tillhandahåller tjänster och stöd till forskare under olika delar av forskningsdatalivscykeln. På grund av de oklara mandat och uppdrag som råder nationellt, så finns förväntningar på de olika aktörerna som inte går att uppfylla. Ett exempel är lagring av forskningsdata, såväl aktiva som inaktiva. Svenska forskare är del av ett globalt forskarsamhälle, och tar därigenom del av en mängd internationella infrastrukturer, verktyg, och sammanhang. I och med den implementeringsfas som EOSC gått in i, så behöver svensk e-infrastruktur vara interoperabel med de internationellt tillgängliga tjänster som svenska forskare använder.

Exakt vilka behov som finns hos svenska forskare, och vem som har ansvar för att svara upp mot behoven, är oklart i många frågor. Det infrastrukturella stöd som forskare behöver för att hantera forskningsdata under hela datalivscykeln, med välfungerande gränssnitt mellan nationella och internationella aktörers e-infrastrukturlösningar finns i vissa fall och är i andra fall fragmenterat. För att komma vidare i utvecklingen av sammanhållet stöd för forskare, så har ett nationellt samarbete startat inom verksamhetsarkitektur för forskningsdataområdet, med syfte att ta fram ett gemensamt bildspråk, visualisera ett mycket komplext område, och att konkretisera krav, villkor, juridiska ramverk, styrdokument etc. som påverkar forskningsdatahantering.

Detta projekt fokuserar på vad forskare behöver ha för att forskningsdata ska hanteras på ett högkvalitativt sätt och vilket infrastrukturellt stöd forskarna i detta. Projektet kommer att fokusera på helheten i datahanteringen, vad som finns på plats på nationell nivå och vad som behöver åtgärdas. Angreppssättet är verksamhetsarkitektur, som verktyg används Vintergatan. En grundkarta för forskningsdataområdet kommer att ligga till grund för projektet.



Figur 1 Grundkarta för forskningsdataområdet

Syfte

Projektets syfte är att föreslå en arkitektur för forskares lagring av forskningsdata, inklusive lokala lösningar och gränssnitt gentemot externa lösningar. Detta innebär att kartlägga forskares behov av infrastrukturellt stöd för forskningsdatahantering under ett forskningsprojekt, med särskilt fokus på lagring av forskningsdata. Projektet kommer att bidra till ökad förståelse för forskningens och lärosätets behov i dagsläget, och peka på områden som behöver utvecklas för att tillhandahålla ett optimalt stöd för forskningsdatahantering.

Mål

Effektmål

Direkt efter projektet: De nationella e-infrastrukturerna och ingående lärosäten har en bild av vilket infrastrukturellt stöd som behövs för forskare gällande lagring av forskningsdata under olika delar av forskningsdatalivscykeln, inkluderande kopplingar till relevant stöd för t ex delning och publicering av data. Detta kan utgöra ett underlag för kravspecifikation för olika delar av det infrastrukturella stödet hos de nationella infrastrukturerna samt hos lärosätena, inklusive lagringstjänster, överföringstjänster, och APIer.

Inom 2 år: De ingående lärosätena har förutsättningarna att utveckla det infrastrukturella stöd som de behöver för att hantera forskningsdata under hela datalivscykeln, inklusive välfungerande gränssnitt mellan lokala, nationella och internationella aktörers e-infrastrukturlösningar.

Leveransmål

Projektet fordrar ett utforskande angreppssätt, som tillåter viss flexibilitet i leveranser, särskilt gällande de kartor och den kravlista som ska levereras, vilka alla utgör levande objekt. Förslag på förändringar kommer att stämmas av med uppdragsgivaren under projektets gång.

Senast den 20 december 2021 är följande levererat:

1. **Kravlista** (compliance list) för lagring, utifrån Chalmers och Örebro universitets organisation och forskningsverksamhet, samt berörda externa aktörers villkor.
2. **Verksamhetskarta för forskningsdatahantering** (en utvecklad version av Figur 1), med särskilt fokus på lagring, inklusive beskrivningar av:
 - Forskarresa: hur forskaren hanterar forskningsdata under hela forskningsdatalivscykeln och därefter, inklusive samarbeten och tillgängliggörande.
 - Vilka aktörer som förekommer (tillhandahåller tjänster) under olika delar av forskningsdatalivscykeln.
 - Gränssnitt mellan aktörer och ägande/ansvar för data, exempelvis när data överförs mellan lärosätet och en annan aktör.
 - Vad forskarna använder i olika delar av forskningsprocessen, exempelvis IT-system, lagringslösningar, e-infrastrukturer, och forskningsinfrastrukturer.
3. **Slutrapport**, innehållandes
 - a) behovsanalys för användarvänlig lagring av forskningsdata under hela forskningsdatalivscykeln, baserat på forskares och lärosätets behov, samt utifrån rådande regelverk,
 - b) designförslag för användarvänlig lagring av forskningsdata under hela forskningsdatalivscykeln, innehållandes användarscenarier och kravspecifikation på funktionsnivå. Designförslaget utgör med fördel grunden till ett införandeprojekt.
 - Rapporten ska publiceras fritt tillgängligt med en bred målgrupp i åtanke.

Tidsplan för slutrapport har högsta prioritering, därefter leveranserna, och sist, resurser.

Projektet är ett samarbete mellan SNIC, SUNET, SND, Chalmers samt Örebro universitet, och resultaten får fritt användas av dessa och andra intressenter utan begränsningar.

Avgränsningar

Implementering av föreslagen design ligger utanför projektet, men en dialog förs med det IT-orienterade projekt på Chalmers som skaffar sig kunskap om och undersöker möjliga tekniska lösningar parallellt med detta projekt

Djupare dialog med en större mängd aktörer än SNIC, SND och SUNET, såsom forskningsinfrastrukturer, ryms inte i detta projekt, men kan med fördel bli fokus i ett fortsättningsprojekt inom det nationella samarbetet för verksamhetsarkitektur för forskningsdata, i mån av resurser.

Upplägg

Tre avstämningar inplanerade med styrgruppen under hösten. På grund av projektets utforskande natur, så är prioriteringar en stående punkt på agendan.

- Mitten av september: fokus på **1. Kravlista**: fre 17/9 kl 13 Office-zoomen
- Början av november: fokus på **2. Verksamhetskarta för forskningsdatahantering**: bestäms i sept
- Mitten av december: fokus på **3. Slutrapport**: bestäms i sept

Projektgruppen informerar styrgrupp om status för projektet en gång per månad, och vid behov.

Projektet kommer att informera externa intressenter om projektets status genom etablerade kanaler, såsom SND:s IT-forum, SUNET-dagarna, och det nationella verksamhetsarkitekturarbetet.

1) Utredda forskarnas och Chalmers behov, utifrån:

- a) Visionssamtal med Chalmers e-commons ledningsgrupp och med den operativa gruppen inom Chalmers Data Office om visionen kring sammanhållen lagring – vad betyder det? Följsamhet i datahanteringen – juridiskt, organisatoriskt, tekniskt (på högnivå)?
- b) Intervjuer och workshops med forskare vid både Chalmers samt Örebro universitet om deras forskningsdatahantering och behov, med särskilt fokus på lagring. Upplägget har testats på Lunds universitet med goda resultat. Metoden är därmed utarbetad, och arbetet kan snabbt komma i gång. Utöver att ge information till verksamhetsarkitekturen och kartan, så kommer intervjuerna även att peka på behov av stöd, identifiering av eventuella flaskhalsar och inkompatibilitet mellan aktörer och system.
 - i) Ca 10 intervjuer (per lärosäte) □ Analys □ Grafisk presentation av resultaten □ Workshop för konsensuskapande och prioritering av behovsbilden
- c) Kravlista (compliance list) med utgångspunkt i de nationella e-infrastrukturerna, utifrån bland annat styrdokument som påverkar hantering av forskningsdata, juridiska ramverk, samt villkor från forskningsfinansiärer och andra intressenter. En grundlista finns utarbetad på lärosätetsnivå, och kan revideras.

2) Design: Översättning av den kravbild på hög nivå, som utkristalliseras i analysen av 1 a, b och c. Designen kommer att fokusera på funktioner som behövs för att skapa en sammanhållen lagring för att lösa forskarnas vanliga aktiviteter (t.ex. lagra långsiktigt, publicera data, koppla data och artiklar).

Organisation

- **Uppdragsgivare:** Sverker Holmgren
- **Styrgrupp**
 - E-commons ledningsgrupp + gruppen POH, LOI, SH, VN, OO, MP
- **Projektledare:** Monica Lassi, senior rådgivare, SNIC
- **Projektgrupp**
 - Emma-Lisa Hansson, projektledare, LUNARC, Lunds universitet
 - IT-kontakter på Chalmers finns i Chalmers grupp ”IT-lösningar för datalagring” som nämns ovan.
 - Mattias Persson, utredare, Örebro universitet

Budget

Finansiering av uppdraget består av insatser från organisationerna SNIC, SUNET, SND, Örebro och Chalmers e-Commons.