



KAFFE-projektet

Statusrapport, nuläge och framtid

Presentation vid SND:s nätverksträff, 22 oktober 2025



SND

Svensk nationell datatjänst | Göteborgs universitet - Chalmers tekniska högskola - Karolinska Institutet - Kungliga Tekniska högskolan - Lunds universitet - Stockholms universitet - Sveriges lantbruksuniversitet - Umeå universitet - Uppsala universitet

KAFFE-projektet

Kartläggning av förmågor inom forskningens e-infrastruktursystem



Snabba KAFFE-fakta

Syfte

- kartlägga hur forskarnas dataanknutna praktiker och behov förhåller sig till existerande e-infrastrukturer
- presentera hur e-infrastrukturlandskapet skulle behöva se ut för att bättre samla och möta behoven i framtiden.

Projektet betraktar *alla* delar av e-infrastruktur, inklusive tekniska lösningar, organisation och mänsklig kompetens, och sett i ett större nationellt perspektiv, inte enbart inom SND.

Angreppssätt

- Bygger på tidigare arbeten lokalt, nationellt och internationellt.
- Beskriver nuläge, behov och glapp.
- Samverkan är centralt
- Utvecklingsprojekt, ej forskning

Resultat: Ett underlag för utveckling av stöd för, och samordning av, forskarnas praktiker nu och i framtiden.

Tidsramar

- **Fas 1** okt 24-feb 25: Syftesformulering, formering projektgrupp, planering
- **Fas 2** feb-maj 25: Uppstart av arbetsgrupper
- **Fas 3** maj-nov 25: Insamling, analys, leveranser

Bemanning och organisation

Projektägare: SND:s styrgrupp

Projektledare: Monica Lassi (konsult, Lassi Consulting)

Projektgrupp:

- Emma-Lisa Hansson (LUNARC, LU, via NeIC, t.o.m. juli 2025)
- Gustav Nilsonne (KI, senior rådgivare SND)
- Karin Westin Tikkanen (senior rådgivare SND)
- Lisa Isaksson (språksamordnare SND)
- Madeleine Dutoit (MAU; utbildningssamordnare SND)
- Stefan Ekman (senior rådgivare SND)
- Sverker Holmgren (föreståndare Chalmers e-commons; SND styrgrupp)

Arbetsgrupper för varje spår

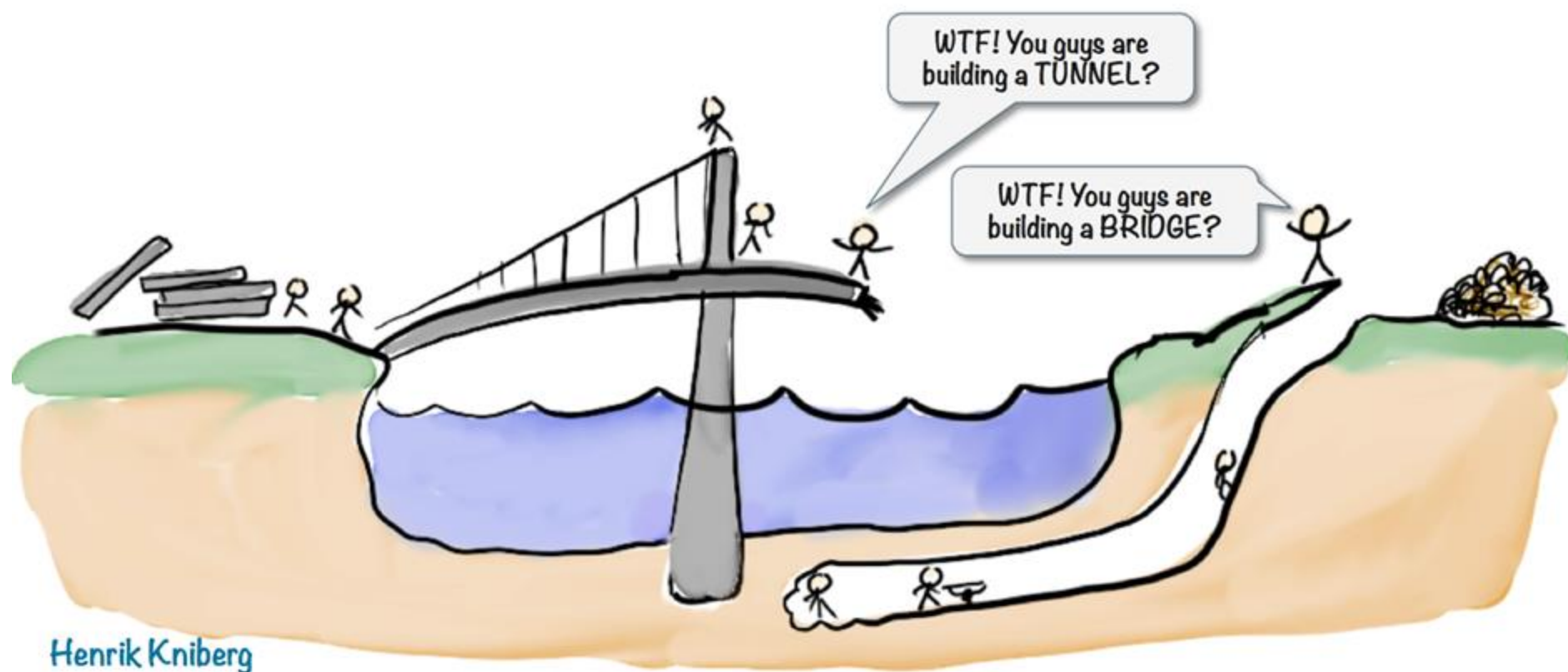
Arbetsutskott med representanter från styrgruppen (Hans Karlsson, Cecilia Björkdahl och Johan Rung) och SND:s ledning.

Referensgrupp med representanter från SND:s konsortium

Samordning - av vad, för vem, av vem?

*Vad betyder samordning,
lagring, forskningsdata,
kund/användare... ..?*

Misalignment

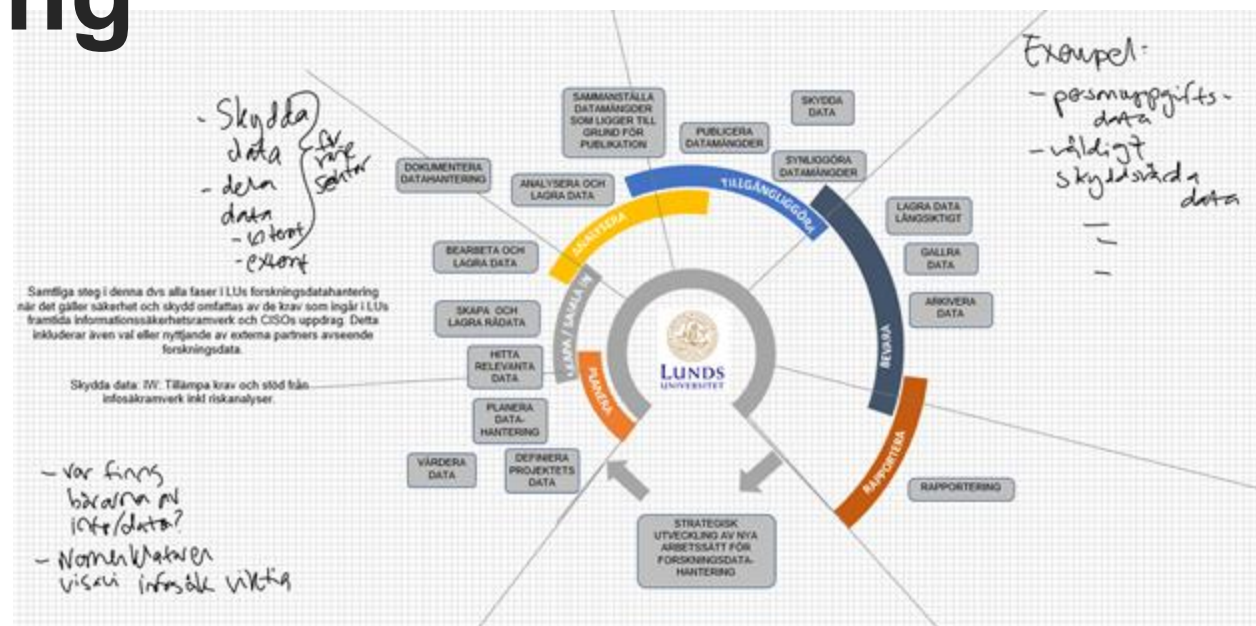


Förmågekartläggning

Förmåga

(verksamhetsarkitektur):

- sammanhållen, avgränsad, självständig komponent av verksamheten
- Fokus på **vad** som ska göras (inte **hur**)
- består av processer, information, kompetenser, regelverk, IT-stöd m.m.

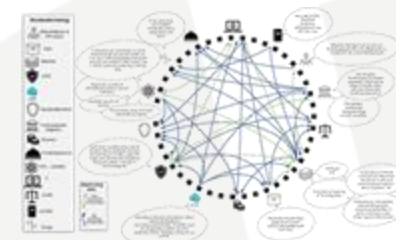


Exempel på förmågor:

- planera datahantering
- skapa och lagra data
- analysera data
- skydda data
- göra data tillgängliga

Angreppssätt och arbetssätt

- Projektet har ett utforskande angreppssätt - flexibilitet utifrån ständigt förändrade förutsättningar i ekosystemet
 - KAFFE lägger grunden till vidare arbete
- Arbetet har organiserats i tre spår som tillsammans skapar en helhet:
 - 1: Analys av rapporter
 - 2: Terminologi och förmågor
 - 3: Intervjuer med forskare
- I slutrapporten vävs resultaten från de tre spåren samman till en helhet.



Resultat

Pågående

- uppdaterad förmågekarta
- uppdaterade beskrivningar av varje förmåga
- beskrivning av områdets terminologi
- uppdatering av forskares nuvarande praktiker och behov
- analyser av glappen mellan forskares behov och stödet som finns idag
- arbetssätt för att samla, förstå, beskriva forskarnas praktiker och behov, för kontinuerlig förankring i forskarnas verklighet

→ Sammantaget: en skattkista av insamlat material, erfarenheter och kompetenser att fortsätta arbeta med!

Framtida möjligheter - exempel

- Fortsatta analyser av allt insamlat material
- Visualiseringar av analyser (kartor och annat)
- Fortsatt insamling av material (alla välkomna att göra intervjuer!)
- Beskrivningar och guider för att systematiskt arbeta vidare på lärosäten
- Gemensam begreppsapparat
- Beskrivningar av nuläge och önskat läge per förmåga
- Översättningar av resultat för att bidra internationellt



Vad säger forskarna?

Ansvariga från KAFFE:s projektgrupp: Madeleine Dutoit och Sverker Holmgren

Tidsram: februari-november 2025.

Syfte: att – med en startpunkt i metodik, guider och svar från tidigare forskarfokuserade analyser – genomföra djupgående intervjuer som bidrar med förståelse för forskarnas praktiker och behov. Ge stöd för de aktörer som ansvarar för att ge det nödvändiga stödet till forskningen och för att samordna olika initiativ (lokalt, nationell, internationellt).

- I arbetsgruppen ingår i nuläget 11 deltagare anmälda från Högskolan i Borås, Chalmers, Handelshögskolan Stockholm, Uppsala universitet, Sveriges lantbruksuniversitet, SND + många antecknare – varmt tack till alla!!
- Dialog förs med KI och GU som genomför eller kommer att genomföra intervjuer.

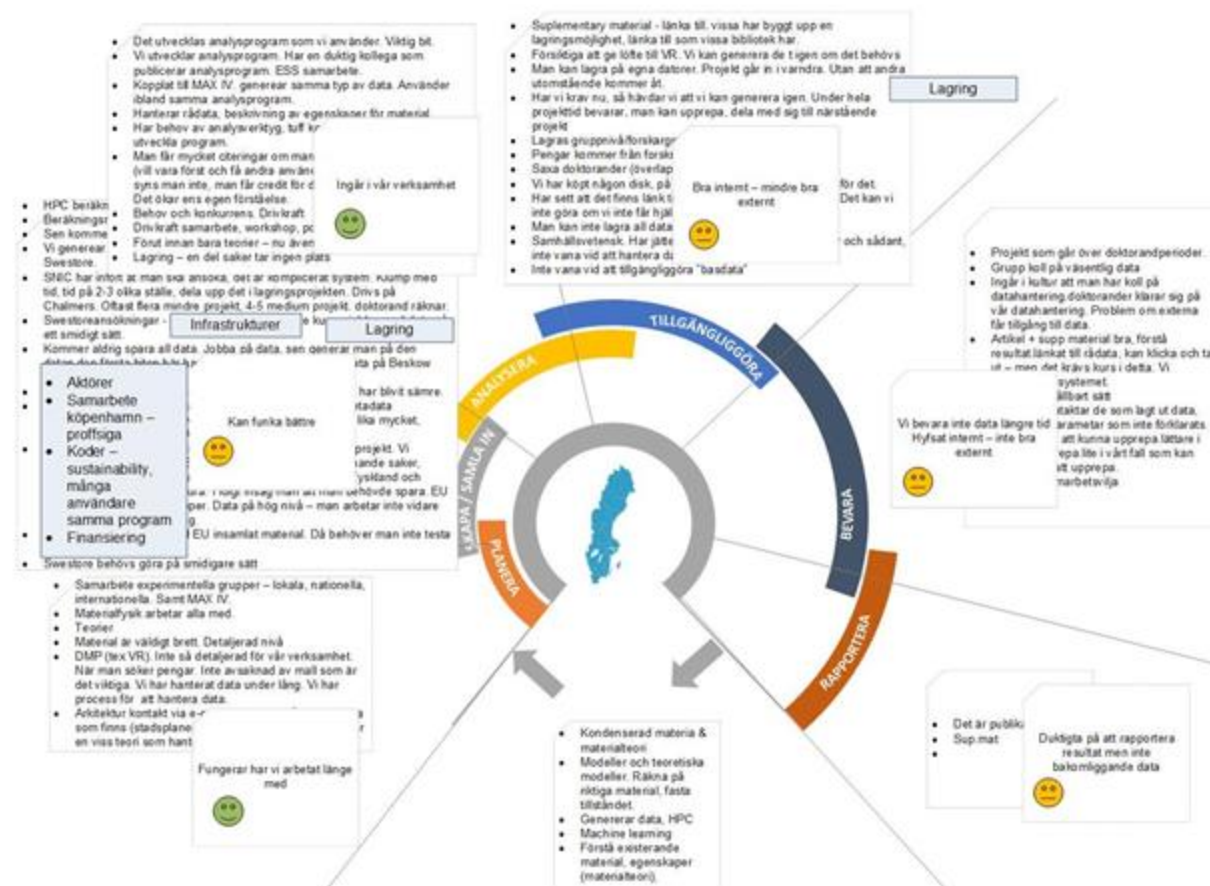
Metod datainsamling

Förmågor och e-infrastruktur centrala begrepp
Forskarnas perspektiv i centrum, deras
praktiker och behov

Data: intervjuer och kartor
39 intervjuer har genomförts.
Deltagarna genomförde 1-11 intervjuer var.
Intervjuerna varade ca 35-60 min.

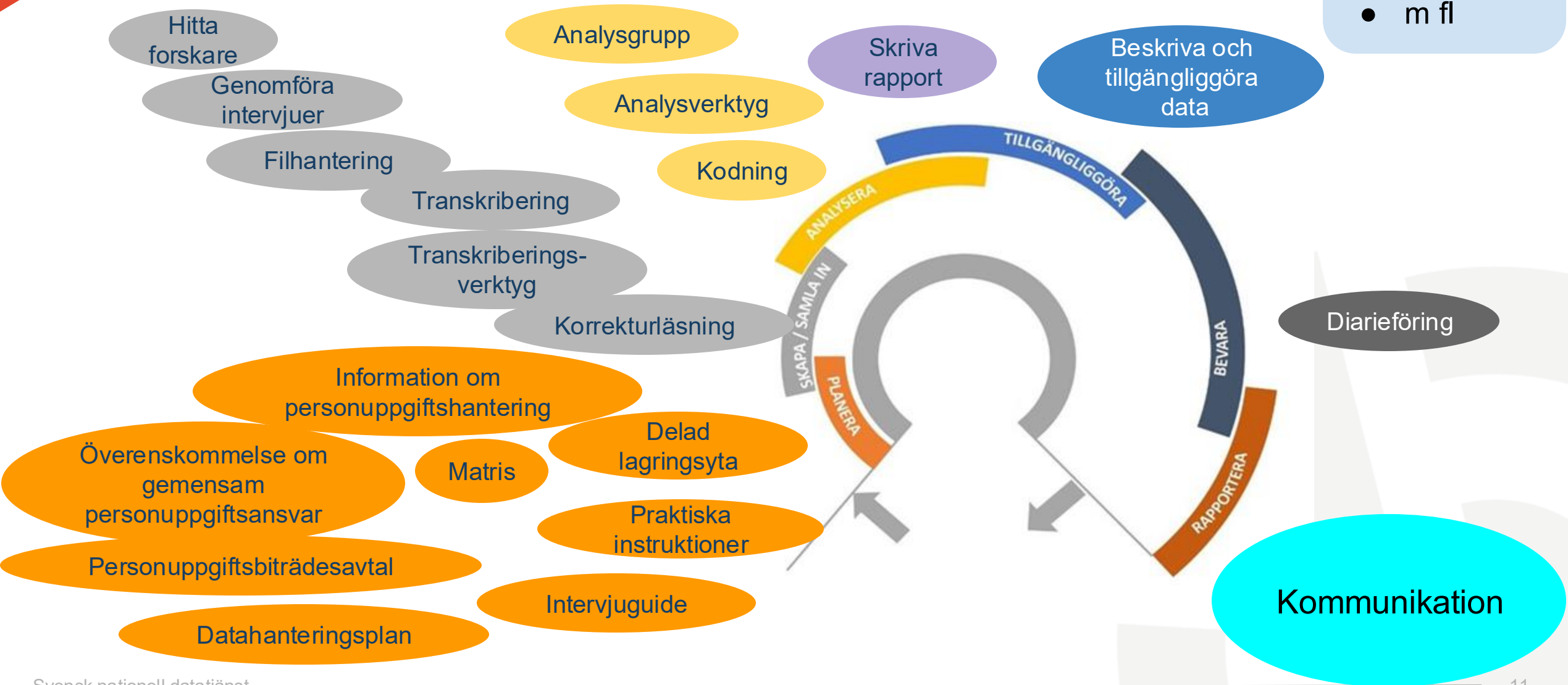
Vi frågade:

- Hur gör du?
- Vad fungerar bra?
- Vad fungerar inte?
- Vilket är ditt drömscenario - hur skulle du önska att det fungerade?



Arbetsprocess

- Jurist
- Arkivarie
- IT
- m fl





Transkribering med AI: metod

- Två öppet tillgängliga AI-modeller för tal-till-text:
 - Svenskt tal: **KB-Whisper large**
 - Engelskt tal: **Whisper large**
- Fri programvara för olika steg i arbetsflödet:
 - **FFmpeg** – konvertera ljud till 16 kHz mono
 - **whisper.cpp** – frontendverktyg för Whispermodellerna
 - **Pandoc** – skapa Word-filer av textutmatningen
 - **Bash** – skapa skript som kör verktygen i rätt ordning



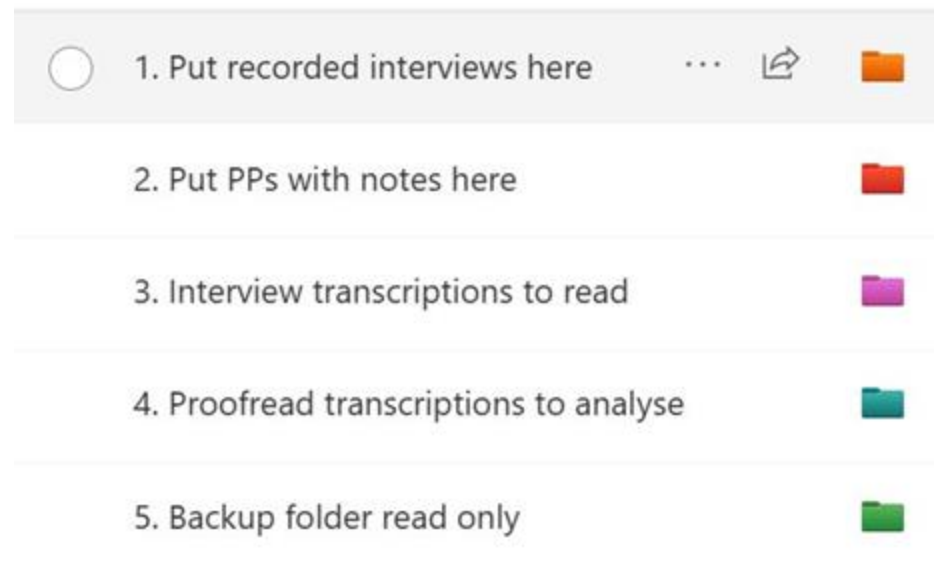
Transkribering med AI: metod

- Inmatning:
 - .m4a eller .mp3-ljud från intervju
- Utmatning:
 - Word .docx och textlogg
- Reflektioner om transkriberingsmetoden:
 - (KB-)Whisper på egen dator går snabbt!
 - 1 timmes tal = 7-9 minuter med GPU-acceleration (Apple M1 Pro från 2021)
 - Modellernas ursprungssyfte – skapa videotextning
 - Talaruppdelning (*speaker diarisation*) – inte en del av modellerna
 - Upphugget tal, störningar – modellerna gissar sämre

Transkribering med AI: utfall

KB Whisper SV och ENG

- behjälpliga blev begärda,
exportera blev exploatera,
Pieta blev Petra
- Ordbortfall
- Uppdelning av talare fick göras manuellt
- Intervjuerna korrekturlyssnades, anonymiserades,
namn, lärosäte och andra identifikatorer togs bort []



Analys

- Analys: kvalitativ, kvantitativ, mänsklig, automatisk
- Tre personer kodar intervjuerna i NVivo, svenska och engelska.
- 23 intervjuer kodas, en per cell

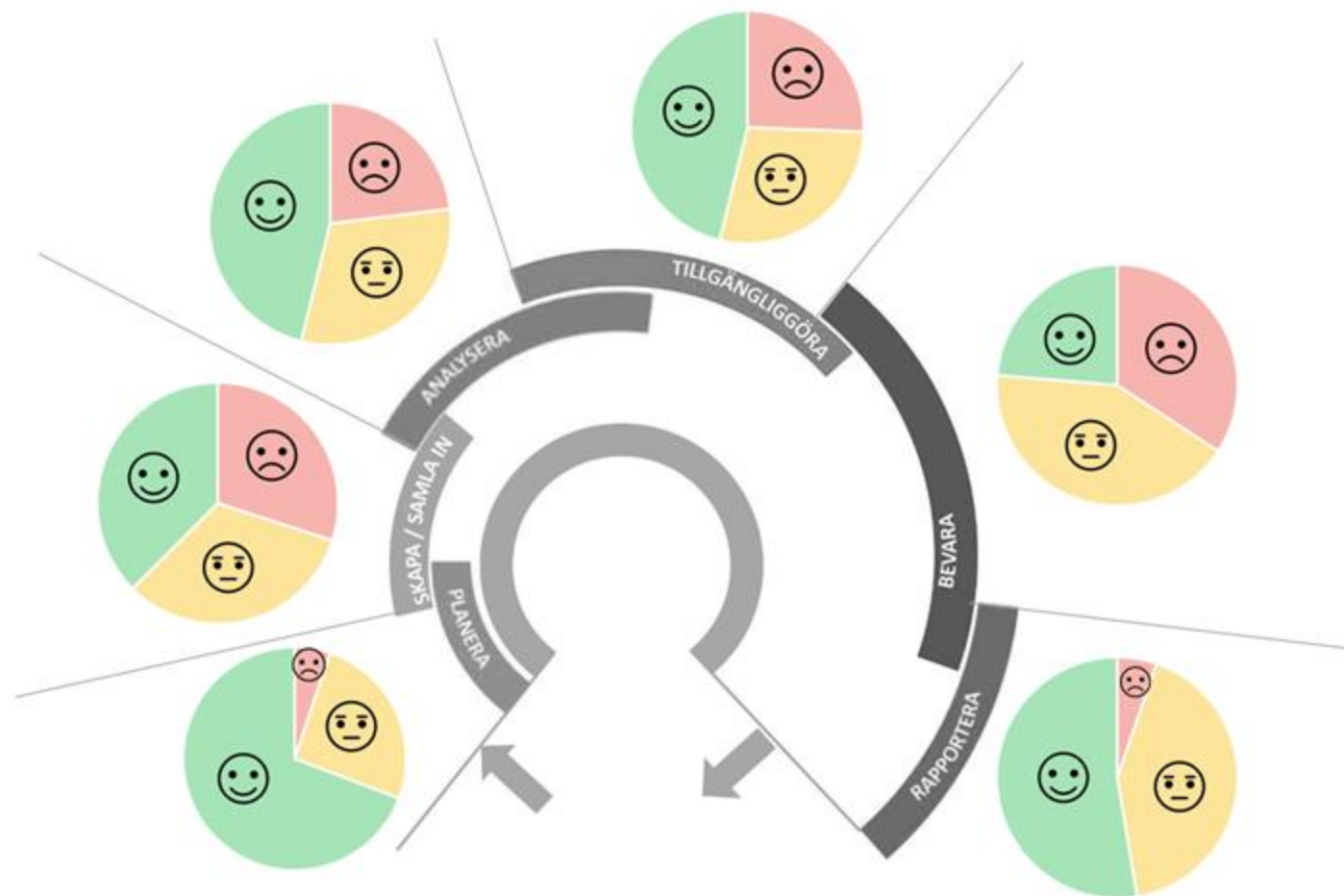
	Doktorand	Postdoc	Junior forskare	Senior forskare (docent/professor)
Humaniora	HumDok		HumJun	HumSen
Konstnärlig forskning	KonstDok			KonstSen
Medicin och hälsa	MedDok		MedJun	MedSen
Naturvetenskap	NatDok	NatPost	NatJun	NatSen
Samhällsvetenskap	SamDok	SamPost	SamJun	SamSen
Teknikvetenskap	TekDok	TekPost	TekJun	TekSen
Lantbruksvetenskap	LantDok		LantJun	LantSen

-
- The diagram is a circular flow chart with five main segments, each representing a stage in the data management process. The segments are labeled in a circular arrangement: 'TILGUNG VON DATEN' (top), 'ANONYMISATION' (left), 'STORAGE' (bottom-left), 'ANALYSE' (bottom-right), and 'BEREINIGUNG' (right). Each segment is associated with a specific challenge or solution, often accompanied by a yellow smiley face icon.
- TILGUNG VON DATEN (Top):**
 - Challenge: Did not make any data accessible, they are stored on a USB, the code key is deleted though. Did not think anyone could have an interest in them let alone use of the data. They are too specific.
 - Solution: Keeps transcripts on USB at home. These are anonymised and code key is deleted.
 - ANONYMISATION (Left):**
 - Challenge: Did not know what tool to use for recording interviews (all online). Used Teams that they knew how to use. Recorded and stored on their computer. Listened, transcribed using KB whisper – that was difficult to access and download. Turned to IT but got no help with the program. A doctoral student helped out.
 - Solution: Did not know where to store data. Used institution storage, home catalogue but was unsure of if this was safe.
 - STORAGE (Bottom-Left):**
 - Challenge: Did a download of NVivo but did not continue, had no time to ask. Did not transcriptions thereby by hand. Did analysis by hand, by photocopying, grouping them, adding notes in margins. Had wished for support from supervisors, information of where to find help, of courses to take.
 - Solution: Planned for anonymising data. Reasoned on tools for collecting data. Focused only on the three first phases, plan, collect, analyse. Was unsure of tool for analysis and transcribing. Did not know where to find support, asked other doctoral students.
 - ANALYSE (Bottom-Right):**
 - Challenge: If the research had not been carried out on distance, there would probably been support and help close at hand. They would probably have used the support then. Had wished for someone to contact. Had wished that information of how to handle and manage data would have been clear. Did not get any info at all of this during doctoral period.
 - Solution: Data have been described in interviews and method chapter. If anyone asks for them they can be presented. Does not know of the legal issues related to deleting data.
 - BEREINIGUNG (Right):**
 - Challenge: Did not make any data accessible, they are stored on a USB, the code key is deleted though. Did not think anyone could have an interest in them let alone use of the data. They are too specific.
 - Solution: Keeps transcripts on USB at home. These are anonymised and code key is deleted.



Nöjdhetsanalys

- Fördelning av nöjdhetsindikatorer per sektor i kartan
- Från de 39 första intervjuerna
- Stärks av kvalitativa analysen



Forskarnas röster: om lagring

“A place for data to go”

Okay, dream scenario - I like it.
It would be perfect if, um, if we had tools in place that help us structuring what we're going to collect, meaning that, **when the collection of data takes place, data has already a place to go** /.../. Yeah, and we would have then this place, this structure already, where data goes. Also, from where data can be just retrieved every time that we need something.
(TekSen)

Nej det var egentligen inte ett bekymmer, för på skolan har vi en datalagringslösning som vi på ett sätt ska använda. Det blir så att **'den här ska du använda, här är din mapp, all forskningsdata ska in där'**. Och på det sättet blir det ganska tydligt.
(HumJun)

Forskarnas röster: om stöd

It will be nice to also have a kind of contact person, a data analyst, data management officer, that works with you from the beginning throughout the project. /.../ So to have a more like structured data management support... So like when we are in analysis.. we don't know what to do, how do we collect this - 'oh let's ask you know someone here can maybe'.. **the same person all the time or at least have a clear.. sort of dedicated... have a point of contact to help navigate.. /.../**

So that's the dream. That would be really helpful.

(SamSen)

“A point of contact”

Forskarnas röster: vad är data?

Deltagare: **Det här tycker jag är svårt när jag håller på med konstnärlig forskning, vad som egentligen är min data.** Jag skapar ju grejer, tredimensionella saker, eller kan vara... också att jag jobbar med text, men oftast tar jag foton /.../

Intervjuare: Vad är ett begrepp som du skulle vilja använda?

Deltagaren: Ja. Alltså material, för det kan ju också vara tyget, eller det kan vara pappret eller vad det nu är jag använder mig av. Data känns lite främmande för mig. Eller jag har inte sett på det som data innan, förrän jag tog den där kursen [hantering av forskningsdata] och började fundera på vad är data inom konstnärlig forskning egentligen? Jag kallar det nog experiment, själva grejen jag gör, men sen vet jag inte om jag har något bra ord för det.

(KonstDok)

Forskarnas röster: datadelning inom projekt

Det som vi försöker använda – speciellt när vi skriver tillsammans med kollegan i det här landet utanför Europa – det är att **det är jättesvårt att dela dokument i SharePoint som är det som vi ska använda**. När det då är en lite konstigt e-postadress, så får inte den personen tillgång till det. Så då vill den personen använda Google Docs och det får vi ju inte egentligen./.../ Så att det jag gjorde nu, det som man behöver göra, för **just den datan./.../ det fick vi ju fysiskt, när vi träffades i det landet**, liksom lägga över från våra maskinen då. Det var verkligen jättesvårt.

(HumSen)

Teams är ju ett sätt som man kan lätt dela med sig av mindre datamängder eller när det är text och när det är dokument och så. Men, jag vet inte hur många Teams-anknytningar jag har, jag tror det är sex eller sju, och det är ju väldigt mycket kollektioner tycker jag när man kommer in i de där. **Eller också är det för att jag är dåligt informerad eller jag är dåligt utbildad. Jag kanske skulle kunna tillägna mig Teams-verktyget på ett bättre sätt**, men i den här övre listen här uppe till höger så tror jag sju olika organisationer som jag ska...

Det är inte alltid det funkar faktiskt.

(KonstSen)



Vad säger rapporterna?

Ansvariga från KAFFE:s projektgrupp: Karin W. Tikkanen och Gustav Nilsonne

Tidsramar: Februari – maj 2025

Syfte: att studera hur de olika utredningarna och rapporterna belyser hela eller delar av det infrastrukturella stöd som krävs för att hantera och dela forskningsdata, med utgångspunkt i förmågekartläggningen.

Spåret uppmärksammar också lokala satsningar på lärosäten eller andra organisationer som syftar till att stärka olika förmågor.

I arbetsgruppen ingick personal från FHS, GU, HHS, KI, MaU, SLU och UU.

	Nämner den valda rapporten den angivna verksamhetsförmågan? (ordagrant eller med omskrivningar)	Om flera sidor är relevanta, ange detta.	Om möjligt, klistra gärna in aktuell mening eller början av stycke så att det blir lättare att jämföra mellan olika rapporter/utredningar.	Hur anger rapporten att den nämnda förmågan kan/bör hanteras? På lärosätet eller med hjälp av en nationell infrastruktur?	Har du/ni någon ytterligare kommentar efter läsningen? Studera särskilt om terminologin/uttryckssätt känns igen från andra (typer av) dokument. Hur beskrivs t.ex. kortsiktig lagring kontra arkivering, etc.
PLANERA	JA/NEJ	Sida/hänvisning	Citat (om lämpligt)	Ansvarsfördelning	Kommentar
Identifiera och definiera projektets data	Ja		Öppet tillgängliga forskningsdata kan på ett 8 lättare sätt hittas, granskas, delas vidare, citeras och återanvändas. Att arbeta för en omställning till öppen tillgång innebär i första hand att bedömningar av 8 möjligheter till öppen tillgång till forskningsdata (inklusive metadata) blir en naturlig del i planering av datahantering.		Tillgången till återanvändbar data bidrar till denna förmåga.
Planera datahantering	Ja				
Värdera data	Nej				Det skulle vara värdefullt om vi kunde utveckla innebörden av denna förmåga
SKAPA & SAMLA IN	JA/NEJ	Sida/hänvisning	Citat (om lämpligt)	Ansvarsfördelning	Kommentar
Hitta relevanta data	Ja		Öppet tillgängliga forskningsdata kan på ett 8 lättare sätt hittas, granskas, delas vidare, citeras och återanvändas. De möjliggör att forskningsresultat kan samlas, lagras, hanteras och tillgängliggöras på ett standardiserat sätt ... 13	VR ska "samordna, följa upp och främja samverkan i arbetet för öppen tillgång till forskningsdata" Lärosätena ansvarar för tekniska lösningar som följer gemensamma standarder. "Forskningsfinansiärer och forskningsutförande organisationer ansvarar för att finansiering av infrastruktur som stödjer öppen vetenskap"	Tillgången till återanvändbar data bidrar till denna förmåga. Nyckelbegrepp är "gemensamma standarder", "tekniska lösningar för att länka samman publikationer", "öppen källkod" och "öppna standarder".
Skapa och lagra rådata	Ja				
Bearbeta och lagra data	Ja		De möjliggör att forskningsresultat kan 13 samlas, lagras, hanteras och tillgängliggöras på ett standardiserat sätt ... Att metoder, protokoll och	Lärosätena ansvarar för stöd, vägledning, och tekniska lösningar.	Nyckelbegrepp är "gemensamma standarder", "tekniska lösningar för att länka samman publikationer", "öppen källkod" och "öppna standarder". Nyckelbegrepp är "öppna licenser" och "interoperabla tekniska lösningar"

Läsningen studerade bl.a. i vilken grad rapporter/utredningar nämner de olika verksamhetsförmågor som ingår i arbetet med forskningsdata.

Läsarna noterade sina observationer i ett standardiserat formulär, med citat och sidangivelser (om relevant), kommentar om ansvarsfördelning (om sådan anges), samt egna kommentarer/reflektioner utifrån läsningen (valfritt).

OBS! Benämningar är inte alltid exakta/ordagranna. Alternativa skrivningar för olika förmågor kan också noteras.

Här: *Nationella riktlinjerna för öppen vetenskap*, läst av Johan Lagerros (KI)

	Krantzutredningen (2021) Uppdrag till VR (2022) VR:s behovsinventering (2022) Indikatorer för öppen ... (2023) Nationella riktlinjerna (2023) Digitalt stöd KTH (2023) VR:s kartläggning (2024) SUHF:s färdplan (2025) VR:s rekommendationer (2025) Öppen tillgång ... (2025)									
PLANERA										
Identifiera och definiera projektets data	2	0	2	0	2	2	0	0	0	0
Planera datahantering	2	2	2	2	2	2	2	2	2	2
Värdera data	0	0	1	0	0	0	2	0	0	0
SKAPA & SAMLA IN										
Hitta relevanta data	2	2	2	1	2	0	0	0	0	0
Skapa och lagra rådata	2	0	2	0	2	2	0	0	0	2
Bearbeta och lagra data	2	1	2	0	2	1	0	0	0	2
Dokumentera datahantering	2	2	2	2	2	2	2	0	0	0
ANALYSERA										
Analysera och lagra data	2	2	2	0	2	2	0	0	0	0
Sammanställa datamängder som ligger till grund för publikation	2	0	0	0	2	0	0	0	0	0
TILLGÄNGLIGGÖRA										
Publicera datamängder	2	2	2	2	2	2	2	2	2	2
Skydda data	2	2	0	2	2	2	2	2	2	2
Synliggöra datamängder	2	2	2	2	1	2	2	0	2	2
BEVARA										
Lagra data långsiktigt	2	0	0	0	2	2	0	1	0	1
Gallra data	0	0	0	0	0	1	0	1	0	1
Arkivera data	1	0	0	0	0	1	1	1	0	1
RAPPORTERA										
Rapportering	0	0	0	0	0	0	1	0	0	0

Samtliga rapporter/utredningar nämner **datahanteringsplaner**.

Förmågor under rubriken **Analysera data** nämns många gånger inte alls.

Samtliga studerade rapporter nämner förmågor under rubriken **Tillgängliggöra data**.

Förmågor som rör arkivering och gallring av data nämns mest i förbigående, ofta med kommentaren "det är svårt att etablera goda rutiner vad gäller forskningsdata".

Endast en rapport nämner förmågan **Rapportering**.

OBS! Tabellen är inte slutgiltig utan kan komma att justeras i samband med den avslutande granskningen.



Ett par slutsatser från Spår 1

- Ingen vet hur mycket forskningsdata som finns, var den finns, eller hur mycket som delas.
- Variationen är stor mellan forskningshuvudmän när det gäller vilket stöd som tillhandahålls och hur det är organiserat.
- Ökad kunskap behövs bland forskare och stödpersonal om t.ex. FAIR och öppna licenser.
- Arkivering hänger inte ihop med forskningsdatahantering (stys av lagstiftning som forskningshuvudmän behöver följa oavsett).



Terminologi: Varför förstår vi inte varandra?

Ansvariga: Stefan Ekman och Lisa Isaksson, SND

Arbetsgrupp med personal från Chalmers, Södertörns högskola och Uppsala universitet



Gnissel i terminologin

- dela/tillgängliggöra/publicera data
- vidareutnyttjande/återanvändning
- lagring/arkivering/bevarande
- informationssäkerhet/datasäkerhet/dataskydd
- god forskningsdatahantering
- öppen-begreppet (vetenskap, tillgång, data, källkod)
- e-infrastruktur/digital infrastruktur
- datakvalitet

Gnissel i terminologin – exemplet forskningsdata

”data som samlas in eller framställs inom ramen för offentligt finansierad forskningsverksamhet. Med data i detta sammanhang avses information i digitalt format.”

”data som till någon del är offentligt finansierade, som samlas in eller framställs inom ramen för vetenskaplig forskningsverksamhet och som görs direkt tillgängliga för vidareutnyttjande genom en dataplattform som är allmänt åtkomlig”

”alla underlag och information, oavsett format, som samlats in eller genererats för att analyseras i ett vetenskapligt syfte. [...]Forskningsdata kan vara både kvantitativa och kvalitativa, och kan utgöras av den primära data som samlats in direkt för ett specifikt forskningsprojekt eller sekundära data som har återanvänts från tidigare studier.”

”andra handlingar i digitalt format än vetenskapliga publikationer, som samlas in eller framställs inom ramen för vetenskaplig forskningsverksamhet och som används som bevis i forskningsprocessen eller som i forskarvärlden är allmänt accepterade som nödvändiga för att validera forskningsrön och forskningsresultat.”

”alla underlag för forskningsresultat. Det kan röra sig om allt från mätningar, statistiska data och intervjuer till dokumentation av en vetenskaplig och/eller konstnärlig process som lett fram till ett forskningsresultat.

”Digital information som har skapats under forskningsprocessen. Om redan befintliga data från andra aktörer har använts, så som exempelvis olika typer av myndighetsdata, utgör dessa forskningsdata endast om de analyserats, bearbetats eller på annat sätt förädlats under forskningsprocessen”

Det här menar SND med forskningsdata

Vi föreslår att SND i sitt arbete utgår från följande definition av begreppet *forskningsdata*:

Forskningsdata är representationer av företeelser som, utifrån vetenskapliga principer, bedöms kunna användas för vetenskaplig analys eller för att på andra sätt ligga till grund för eller validera forskningsresultat. De har sammanställts, samlats in, skapats eller på annat sätt genererats som en del av det vetenskapliga arbetet.

Forskningsdata kan förekomma i digitala eller analoga format och komma från forskning som är finansierad på olika sätt. SND arbetar främst med helt eller delvis offentligt finansierade, digitala forskningsdata och data avsedda för forskning.

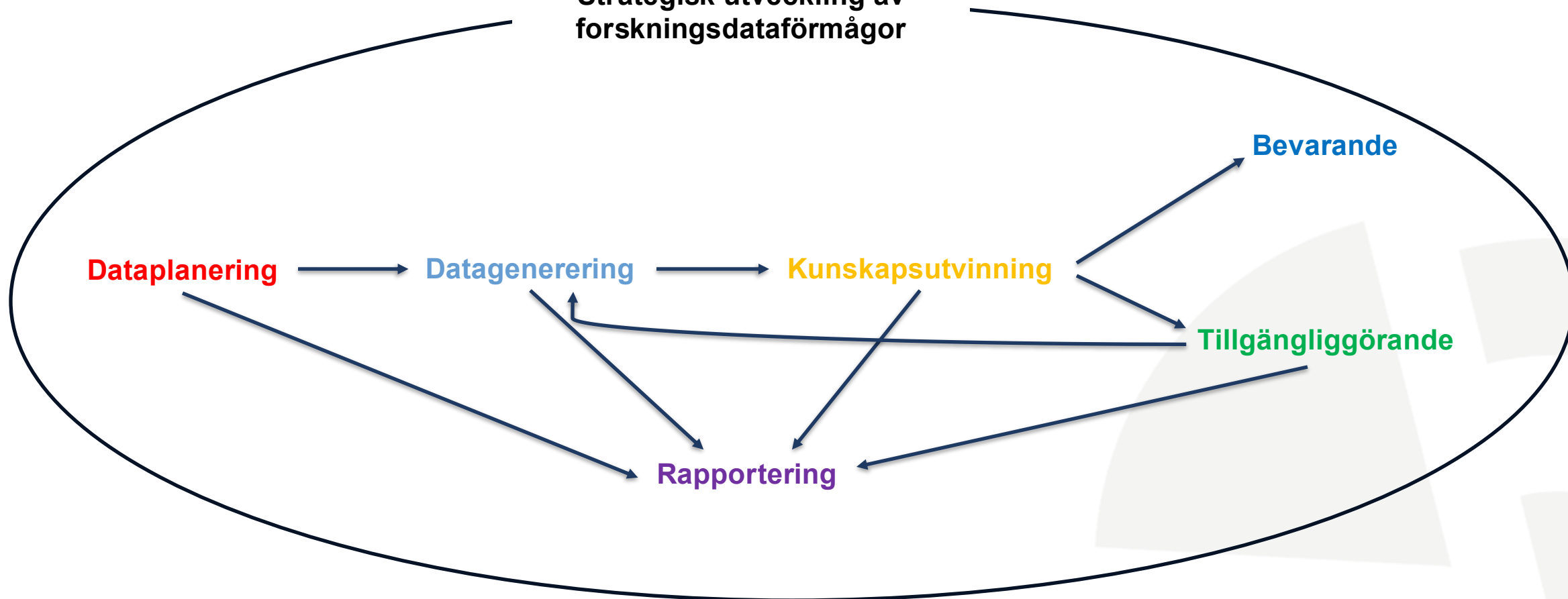


Uppdaterade förmågor - övergripande

- Dataplanering
- Datagenerering
- Kunskapsutvinning
- Tillgängliggörande
- Bevarande
- Rapportering
- Strategisk utveckling av forskningsdataförmågor

Förmågor relaterade till arbetet med forskningsdata

Strategisk utveckling av
forskningsdataförmågor



Uppdaterade förmågor - detaljnivå

Dataplanering

- Datavärdering
- RDM-planering
- Lokalisering av befintliga data

Datagenerering

- Återanvändning av data
- Nyproduktion av data
- Bearbetning av data
- Paketering av data
- Lagring av genererade data
- Dokumentation av data och datahantering

Kunskapsutvinning

- Analys och tolkning
- Lagring för kunskapsutvinning

Bevarande

- Rensning av data
- Slutdokumentation
- Utlämning
- Lagring för långsiktigt bevarande
- Gallring

Tillgängliggörande

- Synliggörande av data
- Åtkomliggörande av data
- Databeskrivning
- Lagring för tillgängliggörande

Rapportering

- Analys och tolkning
- Lagring av information

Strategisk utveckling av forskningsdataförmågor

- Uppföljning och utvärdering
- Utveckling av (nya) förmågor
- Omvärldsbevakning



Fortsättning följer!

- Mycket är påbörjat, mycket kan göras framöver
 - Material och kartläggningar att arbeta vidare med
 - Samverkan på lokal och nationell nivå
 - Europeisk samverkan?
- Några ord från SND:s styrgrupp och ledning