

KI Data Repository – Fas 1, delrapport 1

KI Research Data Office, 2024-04-16



**Karolinska
Institutet**

KI Data Repository – Fas 1, delrapport 1

Innehåll

Sammanfattning.....	6
Introduktion	7
Nuvarande arbetsflöde.....	8
Registrering, publicering och lagring	8
Tillgängliggörande	9
Handlägningsordning för ärenden inkomna till KI via DORIS	10
Projektet KI Data Repository (KI DR)	11
Metod - Tillämpad utvecklingsprocess.....	12
KI DR:s tidsplan	13
Tillämpad lagstiftning och riktlinjer	15
Riktlinjer för Öppen vetenskap.....	15
Lagar, författningar och riktlinjer	16
Varför – Övergripande problem och önskade förbättringar.....	19
Vad – Förmågor och behov	22
Verifiering och autentisering.....	25
Lösningens intressenter	27
Arbetsflöde - Verksamhetsarkitektur.....	31
Administrativa rutiner	41
Ärendehantering.....	41
Spårbarhet i personuppgiftshantering	42
Möjlig teknisk lösning - Lösningsarkitektur.....	43
Logisk systemkarta - översikt	43
Lösningsförslag	45
Datalager	46
Egenutvecklad IT-lösning för identitetskontroll	46
Användargränssnitt och användarvänlighet	47
Diskussion: förslaget arbetsflöde i förhållande till aktuell lagstiftning.....	48

Identifiering av användare	48
Ansaret att bedöma om forskningsdata innehåller personuppgifter	48
Ansaret hos den som lämnar ut forskningsdata.....	49
Etikstillstånd för återanvändning av forskningsdata	50
Diskussion: föreslaget arbetsflöde och säkerhetsaspekter	51
Genomgående informationssäkerhet	51
Verifiering och autentisering av användare	52
Tillgängliggörande av forskningsdata för extern användare.....	53
Referenser	54
Bilaga 1, Grundprinciper verksamhets- och lösningsarkitektur.....	56
Arkitekturens fyra nivåer	56

Tabeller

Tabell 1, Leveranser, Projektplan för KI Data Repository – Fas 1	12
--	----

Figurer

Figur 1, Nuvarande arbetsflöde, figur hämtad ur SND:s DAU-handbok, avsnittet Guide för inkommande dataset (SND 2021b)	8
Figur 2, Syfte, pains och gains – SND Doris.....	20
Figur 3, Förmågor – SND DORIS.....	23
Figur 4, Intressentmodell - SND DORIS	29
Figur 5, Processkarta - SND DORIS	32
Figur 6, Tillgängliggöra forskningsdata.....	34
Figur 7, Göra data sökbar via SND:s forskningsdatakatalog	35
Figur 8, Hämta öppet tillgängliga data med DOI.....	37
Figur 9, Hämta data med restriktioner med DOI.	38
Figur 10, Begära och hämta data med restriktioner, verifiering av frågeställarens identitet.....	39
Figur 11, Möjliggöra hämtning av ej publicerade data med URL/PID.....	40
Figur 12, KI RDO:s onlinemanual på KIBsvar	41
Figur 13, Logisk systemkarta – KI Data Repository	44
Figur 14, Systemkarta - Lösningalternativ - översikt	45
Figur 15, Grundprinciper för verksamhets- och lösningsarkitektur.....	56
Figur 16, Arkitekturs fyra nivåer, detaljerad beskrivning	57

Versionshistorik

Dokumentversion	Gjorda ändringar	Ändrad av	Ändrad datum
1.0	Upprättande av rapport	Lisa Andersson, Mats Andersson, Maia Dexander, Helena Eckerbom, Martina Gidlöf, Tina Harberts, Glenn Haya, Karin Widin	2023-12-11
2.0	Korrigeringar baserat på feedback från projektets styrgrupp, KI:s juridiska avdelning, samt funktioner för IT- och informationssäkerhet	Lisa Andersson, Maia Dexander	2024-02-12
3.0	Korrigeringar baserat på feedback från Johan Fihn, SND	Mats Andersson, Martina Gidlöf, Tina Harberts	2024-04-16

Sammanfattning

Karolinska Institutets (KI:s) forskare har behov av att publicera och lagra forskningsdata¹ på ett säkert och effektivt sätt. Därför har KI initierat projektet *KI Data Repository – Fas 1* som syftar till att skapa en robust teknisk infrastruktur för centralt stöd till KI:s forskare att öppet tillgängliggöra forskningsdata.

Sammanfattningsvis bidrar projektet till verksamheten med:

- Anvisning till gränssnitt för att ladda upp forskningsdata till yta för öppna data respektive till yta för skyddsvärda data
- Lösning för säker hantering av data (lagring och öppning av filer) vid granskning av handläggare på Data Access Unit
- Lagring av forskningsdata på ytor säkerhetsklassade för lagring av öppna data respektive skyddsvärda data
- Möjlighet att generera länkar till öppna forskningsdata för tillgängliggörande via Svensk Nationell Datatjänsts datakatalog
- Förmågan att verifiera och autentisera identitet hos de som begär ut skyddsvärda data
- Metod för utlämnande av skyddsvärda data
- Spårbarhet i utdelning och återtagande av rättigheter att tillgå skyddsvärda forskningsdata

Rapporten presenterar verksamhets- och lösningsarkitektur, kompletterat med en diskussion av föreslaget arbetsflöde i förhållande till aktuell lagstiftning och säkerhetsaspekter.

¹ ”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” (Europaparlamentet och Europeiska unionens råd, 2019)

Introduktion

Karolinska Institutets (KI:s) forskare har behov att publicera och lagra forskningsdata² på ett säkert och effektivt sätt. Ökat tillgängliggörande³ av forskningsdata förbättrar forskningens kvalitet, ökar dess genomslag, minskar dubbelarbete, förhindrar forskningsfusk och främjar innovation. Öppen tillgång⁴ till forskningsdata underlättar även granskning och verifiering av forskningsresultat. Publicering och lagring av forskningsdata regleras av lagstiftning och riktlinjer, vilka introduceras i avsnittet *Tillämpad lagstiftning och riktlinjer*.

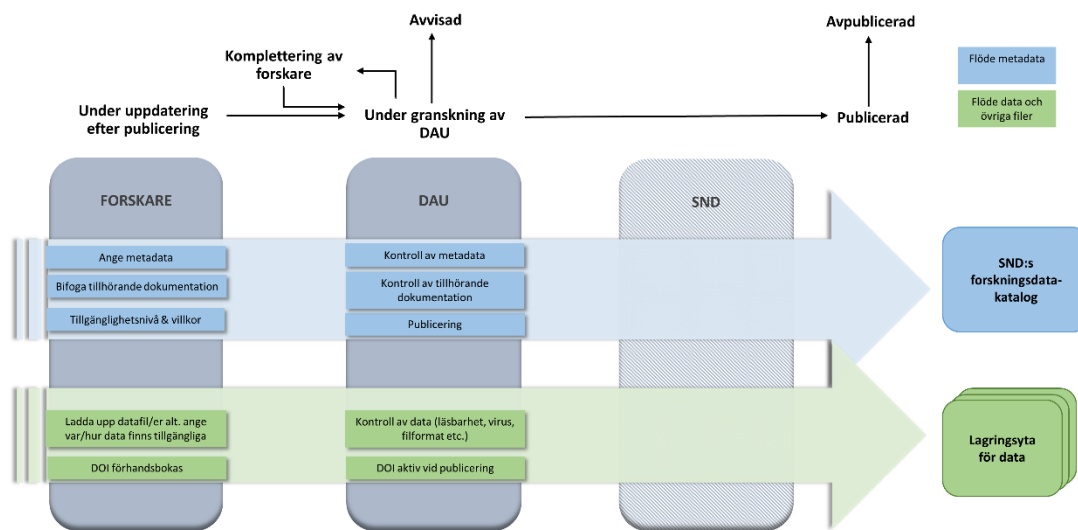
På KI har Research Data Office (RDO) uppdraget att bistå KI:s forskare med ett enkelt och säkert forskningsstöd för i första hand hantering av forskningsdata. Gruppen Data Access Unit (DAU) inom RDO arbetar specifikt för att stödja KI:s forskare att publicera och tillgängliggöra forskningsdata (jfr. SND, 2021a).

² "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" (Europaparlamentet och Europeiska unionens råd, 2019)

³ I rapporten avses att göra forskningsdata så tillgängligt som möjligt med hänsyn till datas innehåll och gällande lagstiftning. "Även om det inte skulle finnas förutsättningar för att tillgängliggöra vissa forskningsdata, kan det finnas möjlighet att göra metadata, det vill säga information om data, tillgängligt." (Vetenskapsrådet, 2022)

⁴ I rapporten avses att ge så öppen tillgång som möjligt med hänsyn till datas innehåll och gällande lagstiftning.

Nuvarande arbetsflöde⁵



Figur 1, Nuvarande arbetsflöde, figur hämtad ur SND:s DAU-handbok, avsnittet Guide för inkommande dataset (SND 2021b)

Registrering, publicering och lagring

- KI:s forskare registrerar sina forskningsdata i DORIS, Svensk Nationell Datatjänsts (SND) dataorganiserings- och informationssystem. Dataset beskrivna och delade via DORIS blir sökbara i SND:s nationella forskningsdatakatalog.
- Data som kan göras öppet tillgängliga laddas upp via DORIS till SND Central Repository (SDN CARE), ett Core Trust Seal (CTS)-certifierat repository. Känsliga forskningsdata överförs via Managed File Transfer (MFT)⁶ till DAU, som sedan för över forskningsdata till S3 via en klient.

⁵ Kompletterande beskrivning finns i *Varför - Övergripande problem och önskade förbättringar*

⁶ Managed File Transfer (MFT) är en tjänst som kan användas för överföring av känsliga forskningsdata från forskarens lagring till DAU.

- En beständig identifierare (Persistent Identifier, PID)⁷ i form av en Digital Object Identifier (DOI)⁸ med KI-prefix reserveras för datasetet hos DataCite, via DORIS.
- DAU granskar metadata⁹ och tillhörande forskningsdata i syfte att förtydliga och förbättra metadatan och forskningsdatans kvalitet, samt att säkerställa att personuppgifter hanteras på ett säkert sätt.
- SND publicerar metadataposten. I och med publiceringen tilldelas datasetet den DOI som tidigare reserverats. Katalogposten i SND:s forskningsdatakatalog blir landningssida för datasetets DOI.

Tillgängliggörande

- Tillgängliggörande av öppet tillgängliga forskningsdata sker genom nedladdning via länkar i SND:s forskningsdatakatalog.
- Öppet tillgängliga dataset lagras på SND:s lagringsyta. Denna lösning är en tillfällig lagringslösning för att dataseten ska kunna vara nedladdningsbara. För effektivt tillgängliggörande av öppna forskningsdata lagrad hos KI behöver forskningsdata kunna göras nedladdningsbar via SND:s datakatalog.
- Det finns två undantag där öppna dataset lagras på KI:s S3-server på grund av att de är för stora för lagring på och nedladdning via SND:s lagringsyta. Åtkomst till dessa två dataset sker efter att användare förfrågar datasetet i SND:s forskningsdatakatalog. DAU besvarar förfrågan via meddelandefunktion i DORIS. DAU skickar en instruktion för hur användaren får tillgång till forskningsdatan.
- Användare får tillgång till öppna dataset som lagras på KI:s S3-server genom att användaren tilldelas ett konto med "key" och "secret key". Key är en publik uppgift, medan secret key motsvarar

⁷ En beständig identifierare (PID) är en unik och beständig digital referens som gör det möjligt att hitta och återanvända digitalt material. PID:ar används för att referera till digitala objekt som dokument, webbsidor och filer. Att forskningsdata är tillgängliga och försedda med en PID är vanligtvis ett krav för att de artiklar som bygger på dessa forskningsdata ska kunna publiceras i tidskrifter.

⁸ Dataset som deponeras och tillgängliggörs via SND förses med persistent identifierare (PID) av typen Digital Object Identifier (DOI). Detta bland annat för att möjliggöra korrekt citering av forskningsdata samt att visa vilken version av forskningsdata som använts.

⁹ I rapporten avses den beskrivande och administrativa information som utgör katalogposter för dataset vilka deponeras i ett forskningsdatarepositorium. Metadata ska vara strukturerade och bör följa vedertagna standarder; "metadata ska, när så är möjligt, vara förenliga med formella öppna standarder." (Europaparlamentet och Europeiska unionens råd, 2019)

ett lösenord. Key och secret key tillsammans ger tillgång till en avgränsad yta på KI:s S3-server där användaren hämtar forskningsdata med hjälp av egenvald klient. Åtkomst via key och secret key utgör inte en tvåfaktorautentisering.

- Det finns ingen rutin för tillgängliggörande av skyddsvärda forskningsdata. Till exempel saknas möjlighet till identitetskontroll av begärande person samt möjlighet att ge åtkomst via tvåfaktorsautentisering.¹⁰

Nuvarande lösning fungerar därför bara delvis som systemstöd i DAU-arbetet. Sammanfattningsvis saknas:

- Lösning för att forskare själva ska kunna ladda upp forskningsdata till KI Data Repository's (KI DR:s) lagringsyta, utan att DAU fungerar som mellansteg.
- Lösning för att lagra öppna forskningsdata hos KI som kan tillgängliggöras genom nedladdning via länk i SND:s forskningsdatakatalog.
- Lösning för utlämning av forskningsdata med begränsad tillgänglighet som lagras på KI DR:s lagringsyta.

Handläggningsordning för ärenden inkomna till KI via DORIS

Det finns två möjliga handläggningsordningar för hur data kan tillgängliggöras efter förfrågan om data inkommen via DORIS:

- a) Data kan tillgängliggöras enligt principerna för utlämnande av offentlig handling.
- b) Data kan tillgängliggöras genom datadelning reglerad av ett Data Access Agreement.¹¹

KI har i nuläget inte beslutat vilken handläggningsordning som kommer att gälla för KI DR. I följande text används begreppet "utlämning", i enlighet med

¹⁰ Förslaget arbetsflöde för att tillgängliggöra forskningsdata med restriktioner finns illustrerat i Figur 9, Hämta data med restriktioner med DOI.

¹¹ En DAA är en sammanfattning av villkoren som gäller för överföring av data. Avtalet beskriver kort vilka data mottagen får tillgång till, vad mottagaren får använda data till och vilka skyddsåtgärder mottagaren ska ha på plats för att skydda data. Ett DAA innebär att personuppgifter överförs från en personuppgiftsansvarig (controller) till en annan personuppgiftsansvarig. Mottagaren av data blir därmed personuppgiftsansvarig.

SND:s bedömning av forskningsdata i SND-katalogen som handlingar som är allmänna (SND, 2022b).

Projektet KI Data Repository (KI DR)

Mot bakgrund av ovanstående har KI behov av ett lärosätesspecifikt data-repositorium där lagring sker lokalt, med rutiner för korrekt utlämnande eller delning. Mot denna bakgrund har KI initierat projektet *KI Data Repository – Fas 1* som syftar till att skapa en robust teknisk infrastruktur för centralt stöd till KI:s forskare att öppet tillgängliggöra forskningsdata (Projektbeställning KI Data Repository Fas 1, 2022).¹²

För att möjliggöra integration med SND:s DORIS och forskningsdatakatalog¹³ i KI DR Fas 2 ska projektet i Fas 1 uppfylla SND:s krav för "Interimslösning för egen lagring utan integration med DORIS" (SND, 2023b).

Projektet är organiserad utifrån sju leveranser:

#	Leverans
L1	Förslag/modell till ett komplett arbetsflöde som omfattar både en möjlig teknisk lösning och de nödvändiga administrativa rutinerna.
L2	Redovisa eventuella konflikter mellan målsättningen att skapa ett välfungerande arbetsflöde för att beskriva och tillgängliggöra forskningsdata med skyddsvärda uppgifter och aktuell lagstiftning, samt säkerhetsaspekter.
L3	Vara till nytta för alla lärosäten inom SND-nätverket som inte ännu har, eller står i begrepp att ordna en lokal lösning för att lagra forskningsdata som ska tillgängliggöras genom SND:s datakatalog. Det på grund av projektet (kring en interimslösning ¹⁴) handlar om förslag till fungerande arbetsflöden, både administrativt och tekniskt.
L4	Fungerande teknisk lösning.

¹² De processer som den tekniska infrastrukturen ska stödja specificeras under rubriken *Arbetsflöde - Verksamhetsarkitektur*.

¹³ För att uppnå projektets mål i Fas 2 krävs integration med DORIS, men också viss integration med datakatalogen då de länkar till öppna data som ska ligga i katalogposterna ska hämta data från lagringsyta på KI.

¹⁴ "Interimslösningen kan användas av organisationer som behöver komma i gång med lokal lagring innan en fungerande integration med SND:s system är på plats... Det är teknisk integration som är slutmålet och det är först då flödet förväntas fungera som planerat. När en integration mellan lokal lagring och SND:s system är på plats ska den permanenta lösningen ersätta interimslösningen." (SND, 2023b, s. 2)

L5	Fungerande arbetssätt, rutiner och processer för DAU-gruppen, forskare och förvaltning.
L6	Överlämning till förvaltning med tydlig ansvarsfördelning.
L7	Utredning för framtagande av plan för långsiktig finansiering av lagring.

Tabell 1, Leveranser, Projektplan för KI Data Repository – Fas 1

Projektet "KI Data Repository – Fas 1" utgör ett av SND:s för närvarande fyra Flaggskeppsprojekt (SND, 2023a) vilket innebär att projektets resultat förväntas vara till nytta för alla lärosäten inom SND-nätverket som inte ännu har, eller står i begrepp att ordna en lokal lösning för att lagra forskningsdata som ska tillgängliggöras genom SND:s forskningsdatakatalog.

Som Flaggskeppsprojekt rapporterar "KI Data Repository – Fas 1" till SND. Föreliggande rapport innehåller:

- (Leverans 1) Ett förslag/modell till ett komplett arbetsflöde som omfattar både en möjlig teknisk lösning och de nödvändiga administrativa rutinerna.
- (Leverans 2) En redovisning av eventuella konflikter mellan målsättningen att skapa ett välfungerande arbetsflöde för att beskriva och tillgängliggöra forskningsdata och behovet av att skydda vissa data enligt aktuell lagstiftning, samt tillhörande säkerhetsaspekter.

Kommande leveranser som rör kommunikation av Flaggskeppsprojektets resultat, teknisk lösning, rutiner och processer, överlämning till förvaltning samt utredning av långsiktig finansiering av lagring rapporteras under 2025.

Metod – Tillämpad utvecklingsprocess

Lösningförslaget som presenteras följer principer för utveckling av verksamhetsarkitektur och lösningsarkitektur (se Bilaga 1), vilket innebär att lösningförslaget beskriver:

- Syfte, pains och gains
- Förmågor
- Intressenter
- Processer
- IT-komponenter, systemkarta lösningsalternativ

Detaljerad design och implementation inklusive framtagande av mjukvaruarkitektur och infrastrukturarkitektur genomförs och rapporteras under 2024.

KI DR:s tidsplan

Projektet är uppdelat i följande steg:

1. Analys
 - a. Syftar till att analysera rådande läge genom att identifiera syfte, förmågor och processer. Analysen ligger till grund för utredningen.
2. Utredning
 - a. Syftar till att ta fram och dokumentera lösningsförslag för projektets styrgrupp att besluta om. Det görs genom att identifiera önskat läge och övergripande design. Vald lösning ligger till grund för kommande implementation.
3. Genomförande
 - a. Syftar till att implementera vald lösning för datalager och identitetskontroll samt att utveckla administrativa rutiner för att leverera ett komplett arbetsflöde.
4. Överlämning
 - a. Syftar till att förbereda projektet för förvaltning, ta fram slutdokumentation samt avsluta Fas 1 som projekt. Här ingår även förberedelser inför Fas 2 (se nedan).

Under projektets gång har tidsplanen justerats för att få med en utökning av projektets leveranser. Den initiala tidsplanen angav att projektet förväntades sträcka sig över tidsperioden Q3 till Q4 2023. Uppdaterad tidsplan sträcker sig i stället till och med Q1 2025.

Projektmålet avgränsas till att uppfylla Fas 1, en interimslösning för egen lagring utan integration med DORIS, men att ge förutsättningar att i nästkommande steg uppfylla Fas 2–3:

- KI Data Repository – Fas 2, permanent lösning inklusive lokal lagring och integration med DORIS enligt *Integrationen mellan lärosätets egen lagring och SND (DORIS)* (SND, 2022a)

- KI Data Repository – Fas 3, exempelvis: utvecklad automatisering av processer och metadataflöden mellan angränsade system som används i forskarstödet vid KI

Tillämpad lagstiftning och riktlinjer

Följande avsnitt redovisar tillämpad lagstiftning och riktlinjer. Inledningsvis presenteras de nationella och lokala riktlinjer för öppen vetenskap som vägleder KI:s arbete med forskningsdata. Därefter presenteras den lagstiftning som reglerar forskningsdatahantering, samt lagstiftning och riktlinjer för IT- och informationssäkerhet som tillämpas i lösningsförslaget. I rapportens avslutande två delar diskuteras presenterat lösningsförslag i förhållande till gällande lagstiftning respektive säkerhetsaspekter.

Riktlinjer för Öppen vetenskap

Projektet *KI Data Repository – Fas 1* vägleds av nationella och lokala riktlinjer för övergång till Öppen vetenskap.

I enlighet med andra principen i Vetenskapsrådets *Vision och vägledande principer för öppen tillgång till forskningsdata* arbetar projektet för att etablera teknisk infrastruktur för öppen tillgång till forskningsdata (Vetenskapsrådet, 2022).

Projektet svarar också mot Rekommendation 4 i Sveriges Universitets- och Högskoleförbunds (SUHF:s) *Vägledning för implementering av färdplan för öppen vetenskap*:

”Att erbjuda forskare prisvärda, adekvata och säkra infrastrukturella tjänster – som uppfyller gällande regelverk (framför allt tryckfrihetsförordningen, offentlighets- och sekretesslagen, arkivlagen och GDPR) och FAIR-principerna – för hantering, lagring, tillgängliggörande och bevarande av forskningsdata och forskningsresultat där arkivering och gallring ingår som en integrerad del i forskningsprocessen och arbetet med öppen tillgång.” (SUHF, 2022)

KI:s *Policy för öppen tillgång till forskningsdata* (2024) anger att:

”KI strävar efter att alla forskningsresultat och forskningsdata ska vara öppet tillgängliga och publiceras tillgängligt på internet i enlighet med FAIR-principerna. Om forskningsdata inte kan göras fritt tillgängliga ska man sträva efter att metadata tillgängliggörs öppet och att forskningsdata sedan tillgängliggörs vid förfrågan efter prövning.” (KI, under beredning)

Lagar, författningar och riktlinjer

Nationella lagar, författningar och lokala riktlinjer för hantering av forskningsdata forskningsdatahantering	Tillämpade principer
Arkivlag (1990:782)	<p>Myndigheternas arkiv ska bevaras, hållas ordnade och vårdas så att de tillgodoser</p> <ol style="list-style-type: none"> 1. rätten att ta del av allmänna handlingar, 2. behovet av information för rättskipningen och förvaltningen, och 3. forskningens behov.
Dataskyddsförordningen (2016/679, GDPR)	<p>För forskning vid universitet och högskolor är det lärosätet, ytterst dess styrelse (motsvarande) som är personuppgiftsansvarig.</p> <p>Den personuppgiftsansvariga ska se till att behandling av personuppgifter sker i enlighet med dataskyddsförordningen. Det handlar till exempel om ändamålet med behandlingen, gallring och säkerhet.</p> <p>Om känsliga personuppgifter behandlas för forskningsändamål måste den personuppgiftsansvarige bland annat vidta lämpliga och särskilda åtgärder för att säkerställa den registrerades grundläggande rättigheter och intressen. Ett exempel på en skyddsåtgärd vid hantering av känsliga personuppgifter är ett obligatoriskt etikgodkännande i enlighet med etikprövningslagen (2003:460).</p> <p>Alla organisationer som behandlar personuppgifter ansvarar för att vidta lämpliga tekniska och organisatoriska säkerhetsåtgärder för att skydda uppgifterna.</p>
Dokumenthanteringsplan för Karolinska Institutets handlingar, Diarienummer 1-673/2023	<p>Verksamhetsområde 4 – Forska</p> <p>Forskningsdata (Process 4.4) kan i vissa fall gallras. Om en handling bedöms ha ett fortsatt inomvetenskapligt värde eller ett värde för annat forskningsområde och bedöms vara av stort vetenskapligt, kulturhistoriskt, eller personhistoriskt värde, eller bedöms vara av stort allmänintresse, får handlingen i fråga undantas från gallring och i stället bevaras. Denna bedömning görs av ansvarig forskare.</p>

Lag (2003:460) om etikprövning av forskning som avser människor	Om forskning bedrivs på levande eller avlidna människor, på biologiskt material från människor, eller involverar hantering av känsliga personuppgifter, krävs etikprövning och godkännande av Etikprövningsmyndigheten.
Offentlighets- och sekretesslag (2009:400) (OSL)	<p>En stor del av de forskningsdata som innehåller personuppgifter och som samlats in av lärosätet, antingen genom egna undersökningar och studier eller från befintliga källor, omfattas av den så kallade forsknings- och statistiksekretessen.</p> <p>Om datasetet omfattas av någon bestämmelse i OSL behöver lärosätet pröva om de ändå kan lämnas ut. I den prövningen kan både uppgifternas art och mottagarens ändamål med användningen av uppgifterna behöva vägas in.</p> <p>Vid utlämnande av forskningsdata via DORIS ska sekretessprövning (prövning av skada och men) genomföras.¹⁵</p>
Tryckfrihetsförordning (1949:105)	<p>Forskningsdata som skapas på statliga myndigheter faller vanligtvis inom ramen för vad som utgör en allmän handling, så som begreppet definieras i Tryckfrihetsförordningen (1949:105).</p> <p>Forskningsdata ska precis som andra allmänna handlingar hanteras i enlighet med reglerna i arkivlagens regler.</p>
Riksarkivets föreskrifter och allmänna råd om gallring av handlingar i statliga myndigheters forskningsverksamhet (RA-FS 1999:1)	<p>Myndigheten får gallra handlingar som inkommit eller upprättats i myndighetens forskningsverksamhet.</p> <p>Gallring skall ske vid tidpunkt eller efter frist som fastställs av myndigheten.</p> <p>Från gallring skall alltid undantas handlingar som innehåller grundläggande uppgifter om syfte, metod och resultat i resp. forskningsprojekt.</p> <p>Handlingar undantas från gallring som bedöms ha ett fortsatt inomvetenskapligt värde eller värde för annat forskningsområde, som bedöms vara av stort vetenskapshistoriskt, kulturhistoriskt eller personhistoriskt värde, eller som bedöms vara av stort allmänt intresse.</p>

¹⁵ Vägledning för att genomföra sekretessbedömning finns i dokumentet "Checklista för utlämnande av forskningsdata med personuppgifter", under bearbetning av SND:s arbetsgrupp Förmedling av forskningsdata med personuppgifter.

Riktlinjer för forskning vid KI, Diarienummer 1-21/2021	<p>Det är viktigt att data beskrivs tydligt, är spårbar, hanteras säkert och att det enkelt går att härleda exempelvis publikationer till underliggande data.</p> <p>Vid hantering av personuppgifter inom forskning så ska data hanteras säkert i enlighet med gällande lagar och regler för dataskydd, samt riktlinjer vid KI för informationssäkerhet och IT-säkerhet.</p>
Riktlinjer för forskningsdokumentation och datahantering vid Karolinska Institutet, Diarienummer: 1- 20/2021	<p>Forskningsdata ska vara så öppna och tillgängliga som möjligt. Detta gäller även andra data som är en del av forskningen, till exempel programkod och skript. Om data innehåller personuppgifter som direkt eller indirekt kan spåras till en nu levande person så får dessa inte göras direkt öppet tillgängliga. För att även kunna ge tillgång till dessa data så publiceras metadata öppet och sedan krävs det reglerad åtkomst till data innehållande personuppgifter.</p>

Lagstiftning och policy för IT- och Informationssäkerhet	Tillämpade principer
KI:s Informationssäkerhetspolicy, Dnr 1-227/2021	<p>KI ska säkerställa att det i alla lägen finns en väl avvägd och ändamålsenlig informationssäkerhetsnivå.</p> <p>Skyddet av informationstillgångar och informationssystem ska vara utformat så att verksamhetens krav på konfidentialitet, tillgänglighet, riktighet och spårbarhet uppfylls.</p>
Myndigheten för samhällsskydd och beredskaps föreskrifter om säkerhetsåtgärder i informationssystem för statliga myndigheter (MSBFS 2020:7)	Myndigheten ska genomföra riskbedömning för enskilda informationssystem och myndighetens produktionsmiljö i sin helhet.

Varför – Övergripande problem och önskade förbättringar

Figur 2 illustrerar KI:s övergripande problem och önskade förbättringar på området forskningsdatahantering. Den önskade förbättringen som lösningsförslaget ska bidra till är att ge forskare kostnadseffektiv och säker hantering, lagring, tillgängliggörande och bevarande av forskningsdata och resultat enligt gällande regelverk och FAIR-principerna¹⁶. Lösningen är specifikt inriktad på att möjliggöra datapublicering via DORIS med lagring på lärosätets lagringyta, samt tillgängliggörande av både öppna och skyddade forskningsdata (SND 2023b). I lösningsförslaget ska öppna forskningsdata tillgängliggöras via länkar i SND:s forskningsdatakatalog. Skyddade forskningsdata ska tillgängliggöras via fildelning administrerad av DAU.

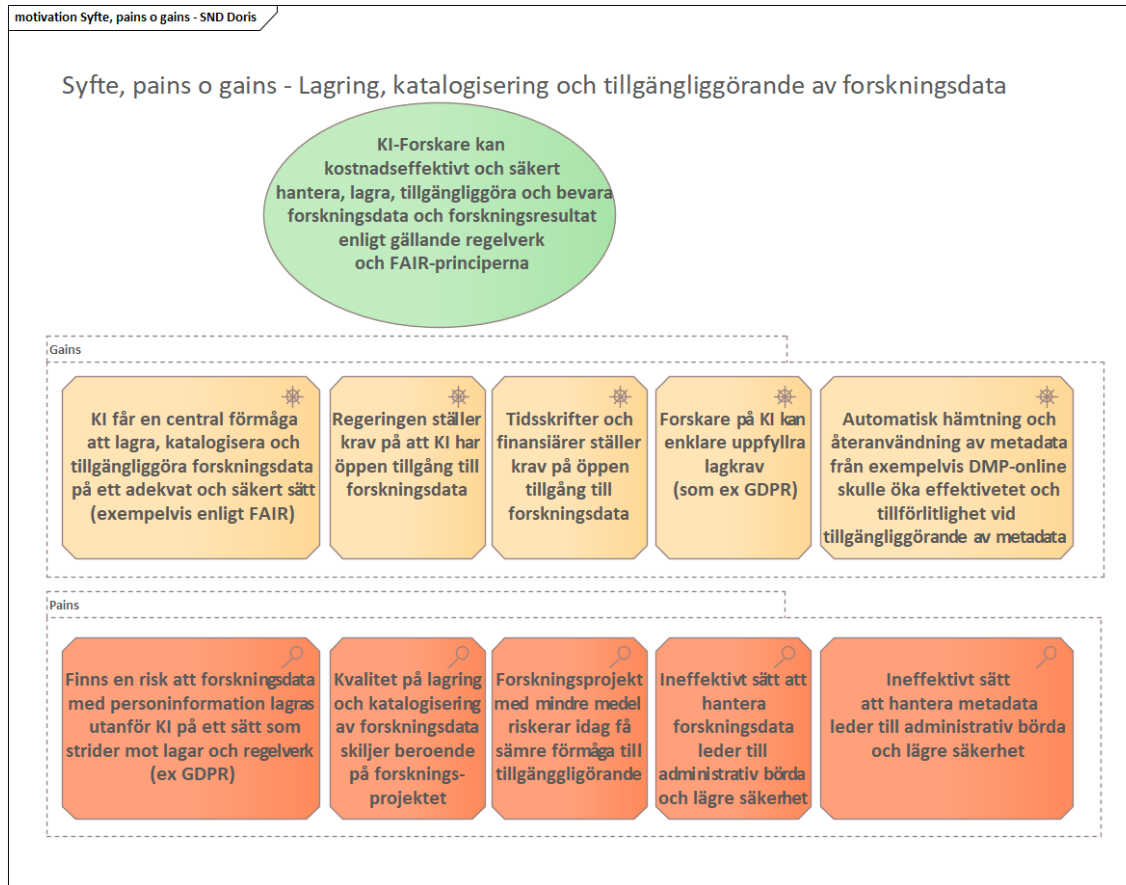
Lösningförslaget ger KI:s forskare tillgång till en säker lagring och utlämning av forskningsdata. Detta är särskilt viktigt då många KI-forskare använder och producerar känsliga personuppgifter. Lösningförslaget innebär också att KI inte längre är beroende av SND för lagring av öppna forskningsdata som ska tillgängliggöras via SND:s forskningsdatakatalog.

Framtagandet av lösningsförslaget omfattar en översiktlig uppskattning av kostnader. Kostnader för investeringar i nya system (till exempel lagring via Sunet Drive) har vägts mot möjligheten att använda redan befintliga system (till exempel lagring på befintlig S3-server). För redan befintliga system finns uppgifter om kostnader för lagring.¹⁷ Lösningförslaget omfattar inte en uttömmande utredning av kort- och långsiktiga kostnader eller finansiering för forskningsdatalagringen. Framtagande av plan för långsiktig finansiering av lagring är en leverans under Fas 1 (se Tabell 1).

¹⁶ FAIR är en akronym som står för Findable, Accessible, Interoperable och Reusable. FAIR-principerna innebär att forskningsdata ska gå att hitta, det ska finnas information om hur man får tillgång till dem, de ska vara kompatibla med andra data, och de ska vara möjliga att återanvända. FAIR-principerna spelar en viktig roll i arbetet för öppen vetenskap och beskriver några av de mest centrala riktlinjerna för god datahantering och öppen tillgång till forskningsdata.

¹⁷ I fallet S3, december 2023: För varje institution inom Samordnad IT ingår 73 TB. Institutionen får fritt fördela volymen på grupp-, projekt-, labbmappar och S3. Volymen som överstiger dessa 73 TB kostar 1,29 SEK/GB/år och interndebiteras kvartalsvis av IT-avdelningen.

Varken lösningsförslaget eller Fas 1 omfattar frågan om automatiserad återanvändning av metadata eller långtidsbevarande av forskningsdata. Utveckling för automatiserad återanvändning av metadata och eventuell integration mot system för långtidsbevarande av forskningsdata kan implementeras under Fas 3.



Figur 2, Syfte, pains och gains – SND DORIS

Figurförklaring

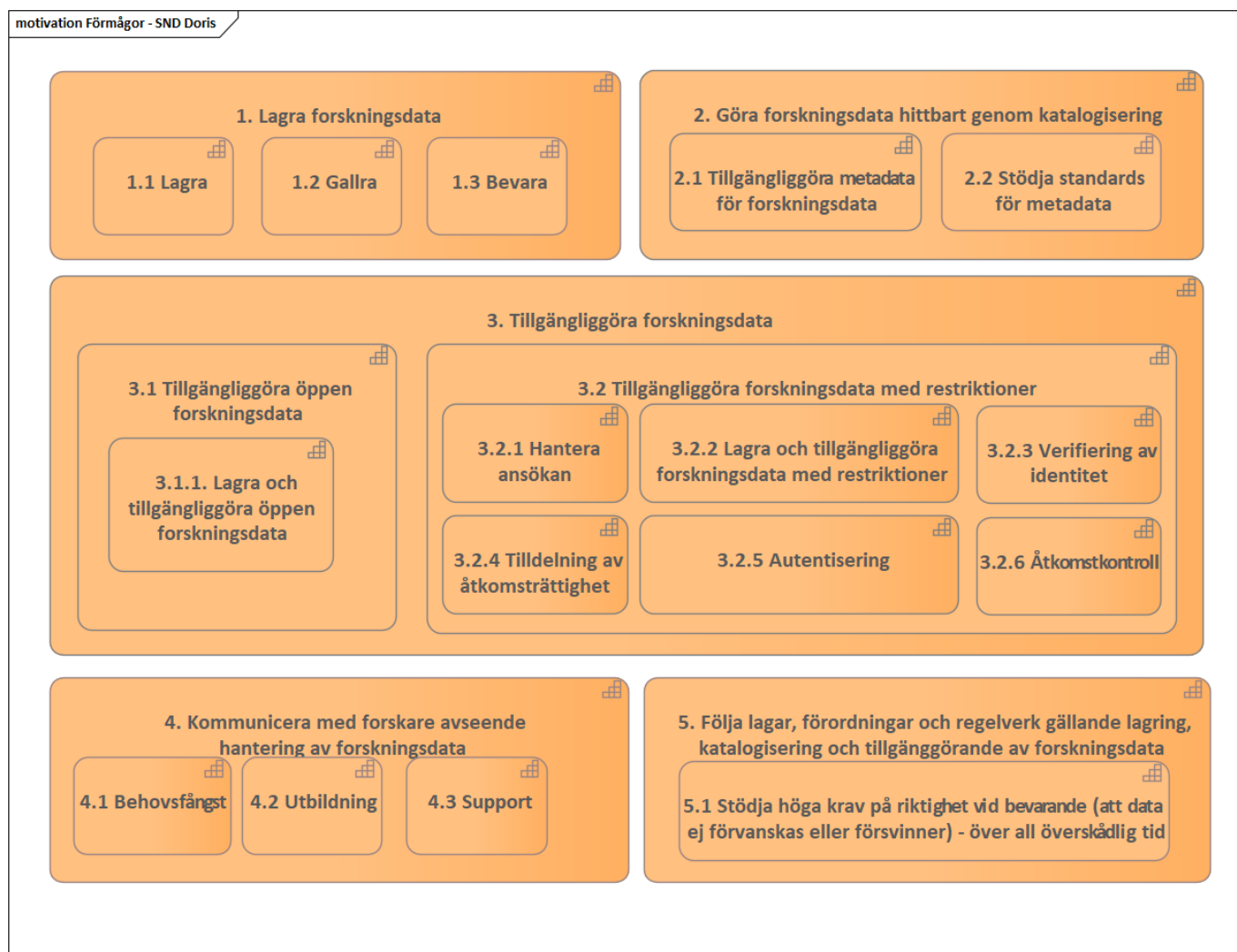
- *Den gröna ellipsen i bilden ovan beskriver "Värdeerbjudande" och är en kort sammanfattning av vad verksamhetsområdet erbjuder.*
- *De beigea objekten med symbolen som ser ut som en skeppsratt beskriver en ambition till förbättring (även kallad "gain" eller "driver").*
- *De orangea objekten med symbolen som ser ut som ett förstoringsglas beskriver en brist (även kallad "pain" eller "assessment").*

Att tänka på när dessa tre typer av objekt identifieras är att de ska ses utifrån och in mot den egna verksamheten, det vill säga utifrån de verksamheter man serverar och samverkar med.

Vad – Förmågor och behov

Lösningsförslaget fokuserar på två av fem förmågor; att göra forskningsdata hittbar genom katalogisering samt att dela och publicera forskningsdata. Fortsättningsvis används begreppet tillgängliggöra forskningsdata som innebär att göra forskningsdata så tillgängligt som möjligt med hänsyn till datans innehåll och gällande lagstiftning. Därutöver innehåller rapporten även en utvärdering av lösningsförslaget i förhållande till lagar, förordningar och regelverk.

Förslaget har därmed inte fokuserat på förmågorna lagring av forskningsdata inklusive arkivering och kommunikation med forskare avseende forskningsdatahantering. I och med att en Core Trust Seal-certifiering är planlagt under Fas 1 kommer dock delar av förmågan långtidslagring utredas och integreras i lösningen. Även planering för kommunikation med forskare ingår i Fas 1.



Figur 3, Förmågor - SND DORIS

Figurförklaring

- *De orangea objekten beskriver verksamhetsförmågor ("business capabilities"). En verksamhetsförmåga kan sägas beskriva ett verksamhets kompetensområde, "vad verksamheten gör och kan".*
- *Verksamhetsförmågorna brukar beskrivas i nivåer. Underliggande förmågor är nödvändiga delar som resulterar i en eftersträvad verksamhetsförmåga.*

Gällande förmåga 2 (se Figur 3), att göra forskningsdata hittbart genom katalogisering, kommer projektet även fortsättningsvis använda DORIS gränssnitt för inmatning av metadata. DORIS utgår från referensmodellen Oberoende Arkivinformatonssystem (OAIS) (ISO 14721:2012, IDT) samt den internationella metadatastandarden DDI (Data Documentation Initiative). Metadata från DORIS är möjligt att exportera i ett antal olika format¹⁸. DORIS stödjer också användande av kontrollerade vokabulärer som SCB:s Standard för svensk indelning av forskningsämnen (SCB, 2011) och ämnesordslistor som Medical Subject Headings (MeSH).

I Fas 1 kommer metadata, precis som idag, läggas in manuellt av forskaren i DORIS. Viss kvalitetskontroll utförs av DAU. För Fas 3 och framåt är det möjligt att systemintegrationer möjliggör återanvändande av tidigare registrerade metadata, till exempel genom hämtning ur ett Research Information Management System (RIMS)–system¹⁹.

Förmåga 3 (se Figur 3), tillgängliggörande av forskningsdata, innefattar både tillgängliggörande av öppna forskningsdata och forskningsdata med restriktioner. För tillgängliggörande av öppna forskningsdata möjliggör lösningsförslaget förmågan lagra och tillgängliggöra öppna forskningsdata. Användare av SND:s katalog kan tillgå öppna forskningsdata via klickbara länkar i SND:s katalogposter som leder till dataset lagrade på KI:s egen lagringsyta. Förmågan innebär att KI inte längre är beroende av SND:s lagring för tillgängliggörande av öppna forskningsdata.

¹⁸ DataCite, DDI 2.5, DDI 3.3, DCAT-AP-SE 2.0, JSON-LD, PDF, Citation (CLS)

¹⁹ KI implementerar för närvarande RIMS-systemet Symplectic Elements.

Gällande tillgängliggörande av forskningsdata med restriktioner innebär lösningsförslaget dels förmåga att *lagra* forskningsdata med restriktioner, dels att *tillgängliggöra* forskningsdata med restriktioner. Tillgängliggörande inleds med att en intressent ansöker om tillstånd att tillgå forskningsdata via förfrågan i SND:s forskningsdatakatalog. Efter inkommen förfrågan prövar DAU, med stöd av arkiv- och juridisk expertis, om forskningsdata går att lämna ut, samt om utlämning kräver sekretessprövning.

Verifiering och autentisering

Innan DAU inleder bedömning av om en person får tillgång till forskningsdata innehållandes känsliga personuppgifter eller andra forskningsdata med restriktioner behöver den sökandes identitet säkerställas. Vid tillgång till forskningsdata via SND:s forskningsdatakatalog sker verifiering och autentisering i KI:s regi, eftersom KI är personuppgiftsansvarig. Minsta acceptabla nivå på säkerställande av sökandes identitet är AL2 (Identity Assurance Level 2²⁰ enligt Sunet, se Sunet 2020).

Vid säkerställandet av identitet på sökande med AL2, för att kunna ge åtkomsträttigheter till data, används olika metoder.

- Verifiering av identitet: I syfte att lämna uppgifter till användaren som används vid autentisering mot en applikation (exempelvis inloggningsuppgifter i form av användar-id och lösenord) behöver en person bevisa att personen är den som personen utger sig för att vara. Detta kallar vi *'verifiera sin identitet'*.
- Autentisering: Vid åtkomst av forskningsdata med restriktioner från KI Data Repository krävs autentisering. Vid autentisering uppger användaren sin identitet genom olika tekniker. Exempelvis genom att uppge användar-id och lösenord. Autentisering kan dock även ske med 2FA (tvåfaktorsautentisering), även kallad MFA

²⁰ Assurance Level – nivån av säkerhet vid verifiering av användares identitet och autentisering

AL 1: kontroll av att personen kontrollerar angiven mailadress (i princip att vi vet att det är en person – men inte vilken person)

AL 2: kontroll av personens identitet på god nivå (till exempel brev till personens folkbokföringsadress med engångskod)

AL 3: kontroll av personens identitet på god nivå och förstärkt autentisering med 2FA

(multifaktorautentisering) då autentisering sker med mer än en faktor. De olika faktorer som typiskt används kan delas in i kategorierna *något man vet* (exempelvis ett lösenord), *något man har* (exempelvis en USB-nyckel) och *något man är* (exempelvis ett fingeravtryck).

- Federation: I en federation, exempelvis SWAMID, litar medlemmarna i federationen på varandras genomförda autentisering. En användare kan därmed genomföra autentiseringen hos en organisation i federationen och ta med sig informationen om godkänd autentisering vid användning av en applikation hos annan organisation i federationen. Därmed behöver inte användaren autentisera sig mot varje organisation, så kallad SSO (Single Sign On). Federationer som SWAMID tillåter alltså användare att autentisera sig en gång hos en medlemsorganisation och sedan använda den godkända autentiseringen för att få tillgång till resurser hos andra medlemsorganisationer utan att behöva genomgå en separat autentisering för varje organisation.
- Åtkomsträttighet och Behörighet: För att få tillgång till data behöver åtkomsträttigheter till data explicit ges till en specifik person. Därmed får användaren behörighet att komma åt data.
- Åtkomstkontroll: Efter att personen autentiserat sig och försöker nå data genomförs åtkomstkontroll då personens behörighet kontrolleras innan den ges tillgång till data.

Lösningens intressenter

Figur 4 illustrerar lösningsförslagets intressenter. Research Data Office (RDO) är beställare av KI Data Repository projektet. RDO är en virtuell organisation på KI bestående av Karolinska Institutets universitetsbibliotek (KIB), Compliance and Data Office (CDO), IT-avdelningen (ITA) och Arkiv och registratur med stöd från Juridiska avdelningen (JA).

Målgruppen för projektet är KI:s forskningsverksamhet. Enligt KI:s delegationsordning är det prefekterna som har ansvaret för institutionernas forskningsdatahantering, inklusive lagring och tillgängliggörande. De primära användarna av lagringslösningen är forskare. Forskningsstödande personal, till exempel bibliotekarier på KIB och handläggare inom CDO, ingår också i användargruppen.

Lösningens övriga intressenter är av olika karaktär:

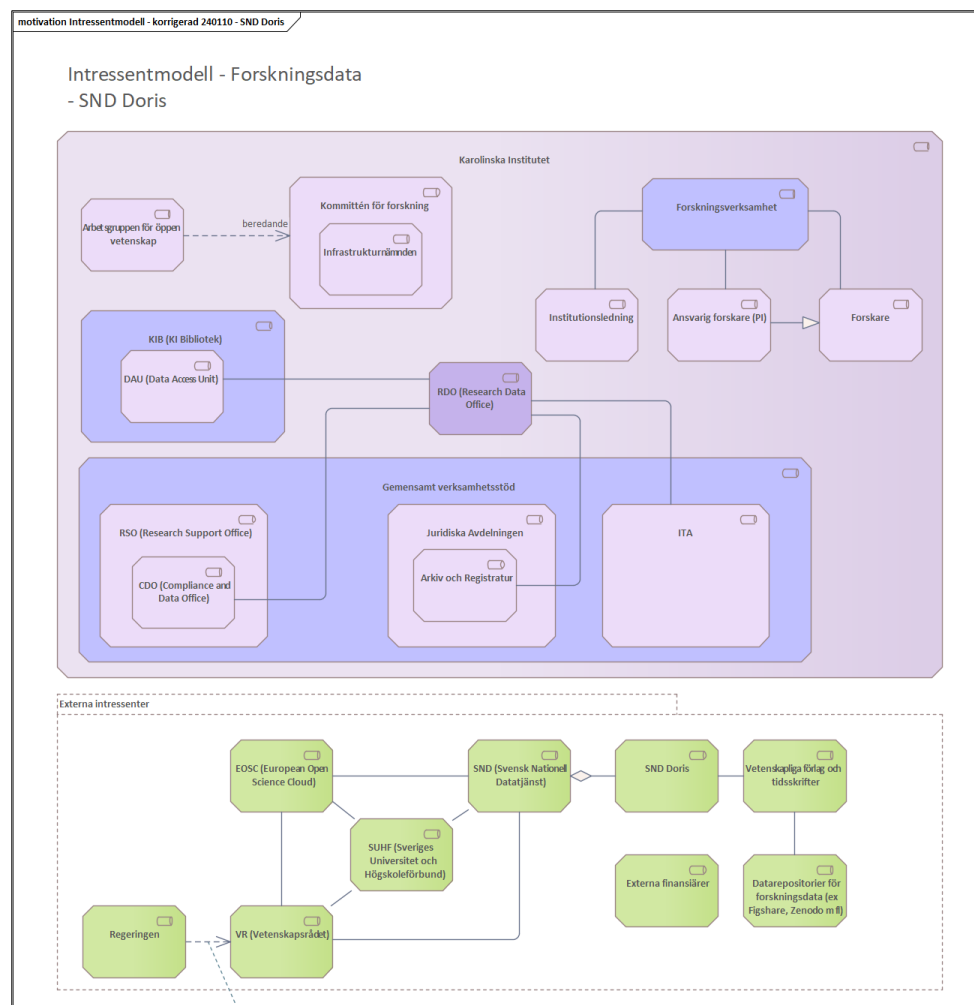
- Policydrivande och beslutsfattande, interna: RDO, Kommittén för forskning, Arbetsgruppen för öppen vetenskap/Open Science Working Group (OSWG), Infrastrukturnämnden
- Policydrivande och beslutsfattande, externa: Regeringen, Vetenskapsrådet (VR), Kungliga biblioteket (KB), Sveriges Universitets och Högskoleförbund (SUHF), Europeiska unionen
- Operativa, interna: RDO
- Operativa, externa: SND

Utöver de intressenter och beroenden som figur 4 illustrerar kommer KI Data Repository förvaltas i en miljö med befintliga och nya relaterade projekt och system. För att effektivisera KI:s infrastrukturella tjänster och bidra till användarvänlig forskningsdatahantering kan integrationer med följande vara relevant i Fas 3:

- KI Open Archive (för pre-prints, parallellpublicering och avhandlingar)
- KI:s e-arkiv (för långtidsbevarande)
- KI RIMS (för sammanhållen forskningsinformationshantering)

- Övriga forskningsstödjande system som tjänster för registrering av kliniska prövningar (till exempel clinicaltrials.gov och Clinical Trials Information System), datahanteringsplaner (till exempel DMP online) och elektroniska loggböcker (till exempel KI ELN)
- Delvis eller helt externa repositorer som SciLifeLab Data Repository, FEGA (The Federated European Genome-phenome Archive) Sweden
- European Open Science Cloud (EOSC)

Karolinska Institutet - KI Data Repository – Fas 1, delrapport 1



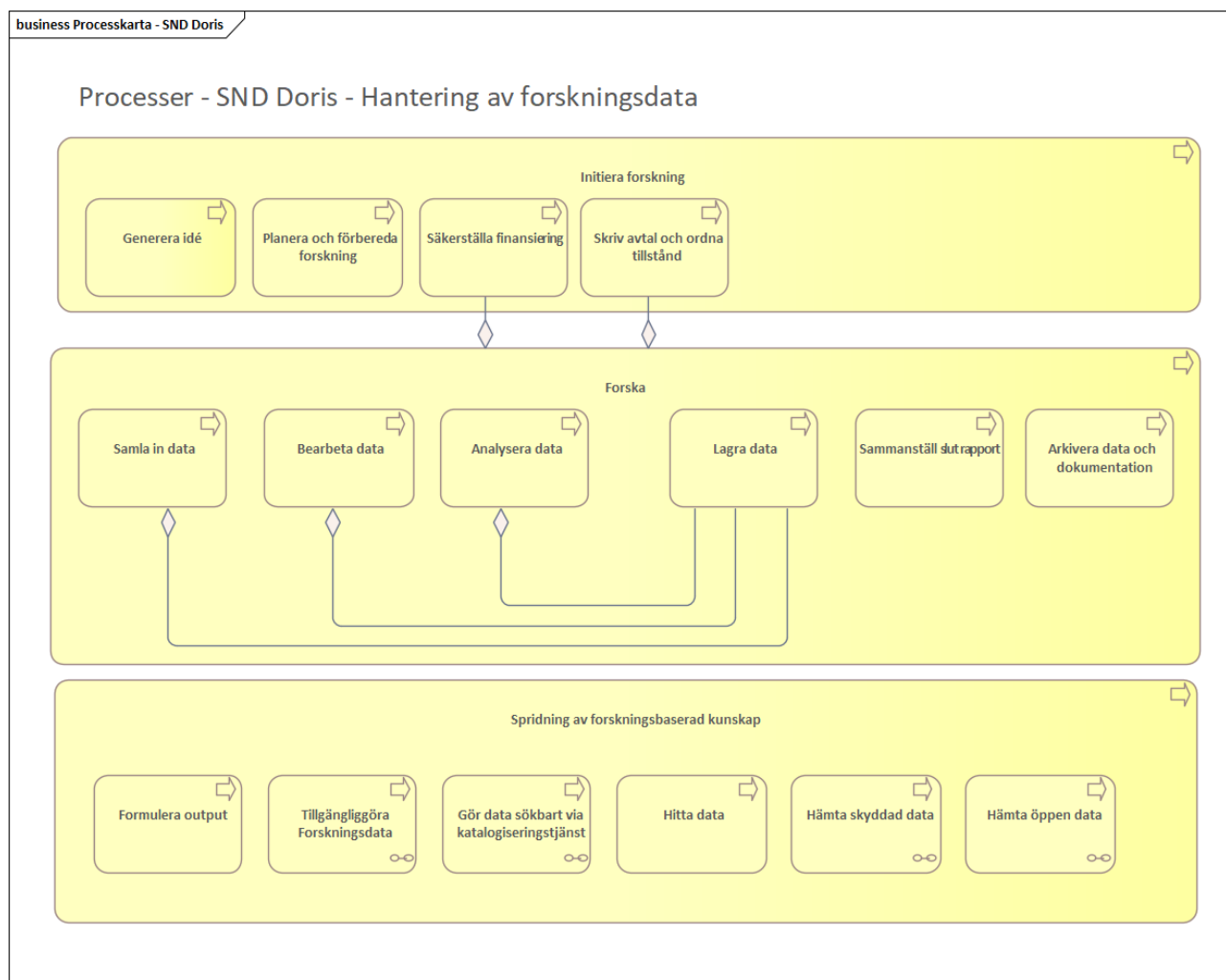
Figur 4, Intressentmodell - SND DORIS

Figurförklaring

- *Objekt med liggande cylinder beskriver en Intressent.*
- *Linjerna mellan objekten beskriver relationer däremellan. Notera skillnaden mellan linjerna. Streckad linje innebär ett informationsflöde. Linje med en "öppen diamant" eller "ruta" beskriver att en intressent innehåller en annan intressent. Exempelvis är ITA en del av UF.*
- *En heldragen linje, utan symbol, beskriver en association.*

Arbetsflöde – Verksamhetsarkitektur

Den inledande figur 5, Hantering av forskningsdata, illustrerar hantering av forskningsdata under forskningsprocessen, samt hur KI Data Repository fyller en funktion för spridning av forskningsbaserad kunskap genom tillgängliggörande av data. Därmed utlämnas andra typer av spridning av forskningsbaserad kunskap, till exempel vetenskapliga publikationer och populärvetenskaplig kommunikation, ur figuren nedan.



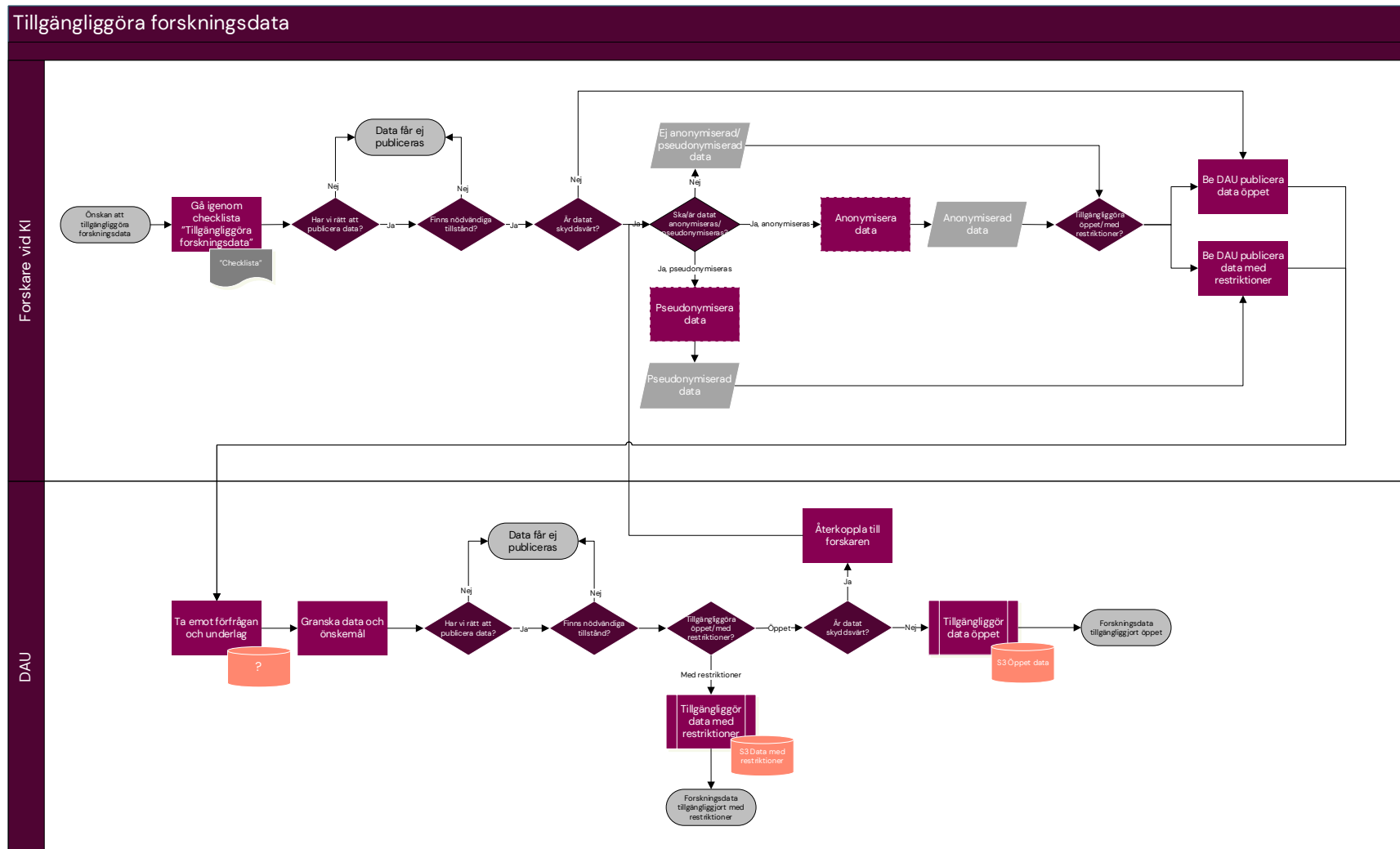
Figur 5, Processkarta - SND DORIS

Lösningsförslaget bidrar till verksamheten med:

- Anvisning till gränssnitt för att ladda upp forskningsdata till yta för öppna data respektive till yta för skyddsvärda data
- Lösning för säker hantering av data (lagring och öppning av filer) vid granskning av handläggare på DAU
- Lagring av forskningsdata på ytor säkerhetsklassade för lagring av öppna data respektive skyddsvärda data
- Möjlighet att generera länkar till öppna forskningsdata för tillgängliggörande via SND:s datakatalog
- Förmågan att verifiera och autentisera identitet hos de som begär ut skyddsvärda data
- Metod för utlämnande av skyddsvärda data
- Spårbarhet i utdelning och återtagande av rättigheter att tillgå skyddsvärda forskningsdata

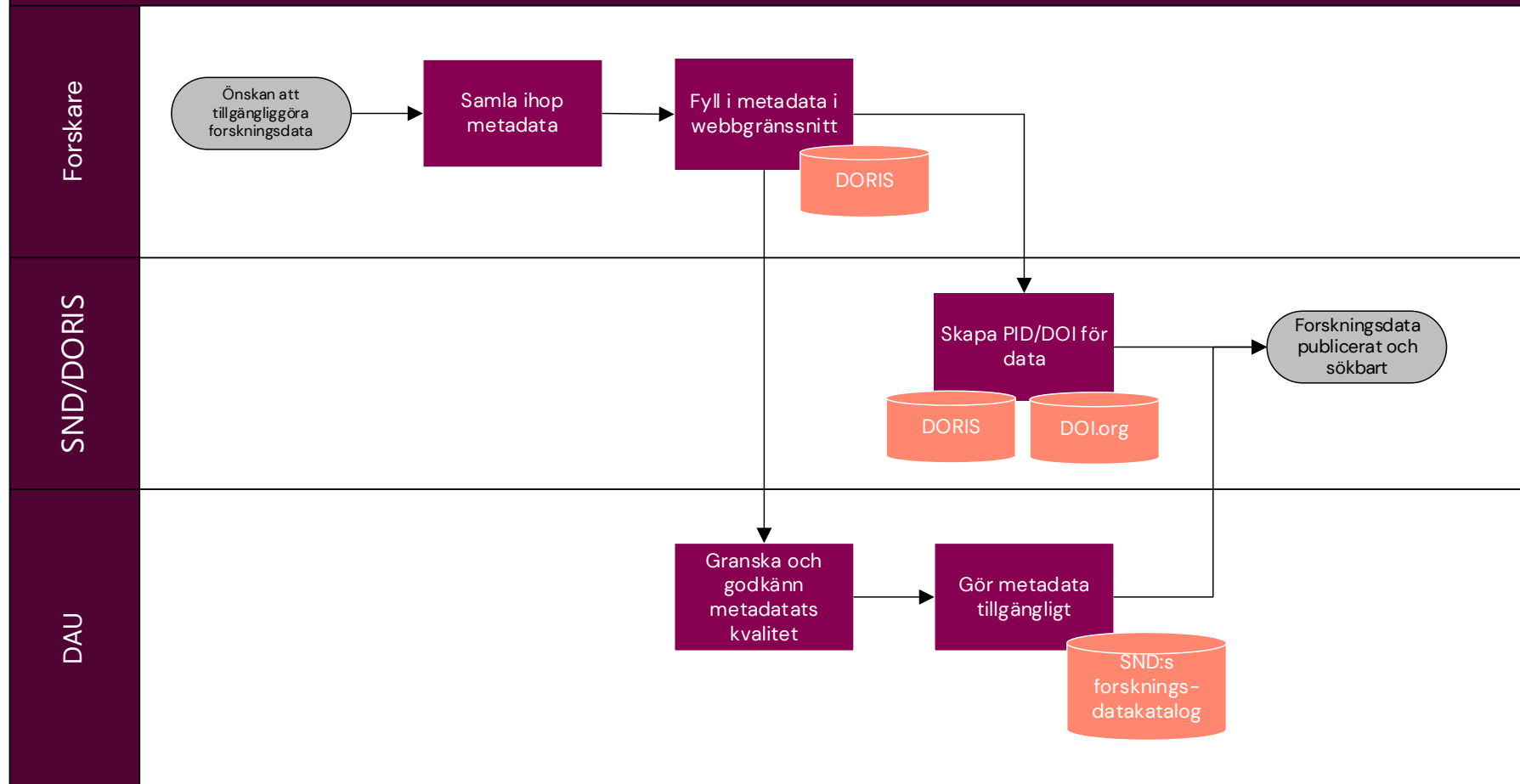
Följande figurer (figur 6–7) illustrerar de aktiviteter som lösningförslaget ska möjliggöra. Figur 6 och 7 illustrerar hur forskningsdata ska lagras för att bli sökbart och kunna tillgängliggöras.

För att tillgängliggöra forskningsdata via SND:s forskningsdatakatalog krävs registrering och granskning av metadata i DORIS. Efter en bedömning av forskningsdatats innehåll beslutas om forskningsdata ska tillgängliggöras öppet, med restriktioner, eller inte tillgängliggöras alls. Vid bedömning tillämpas principer från Offentlighets- och sekretesslagstiftningen samt Dataskyddsförordningen. Forskningsdata tilldelas en PID-DOI. I och med tilldelning av permanent PID-DOI kan forskningsdata sägas vara publicerat.



Figur 6, Tillgängliggöra forskningsdata. Zooma in i PDF-filen för att läsa figurens detaljer.

Göra data sökbart via katalogiseringstjänst – SND/DORIS

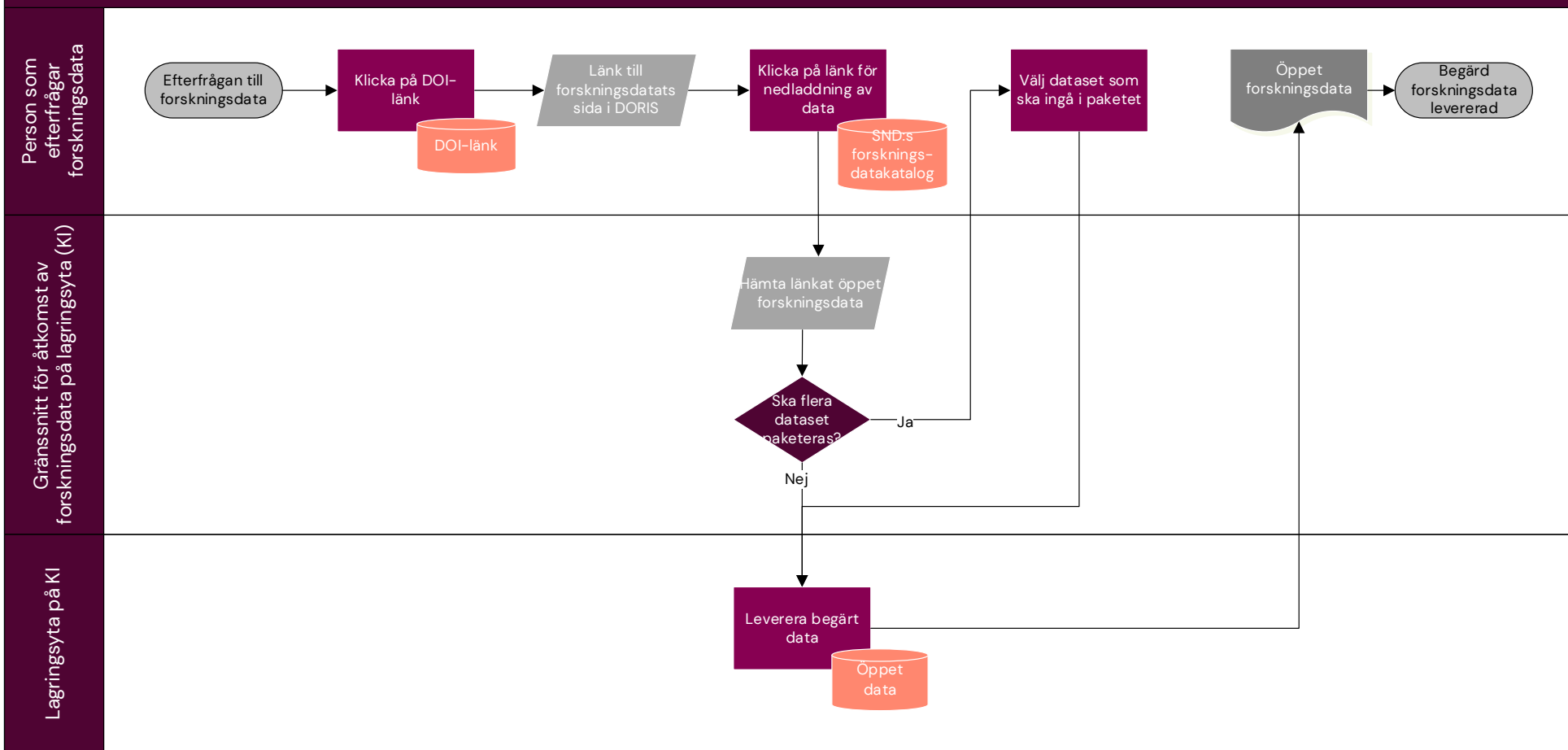


Figur 7, Göra data sökbart via SND:s forskningsdatakatalog. Zooma in i PDF-filen för att läsa figurens detaljer.

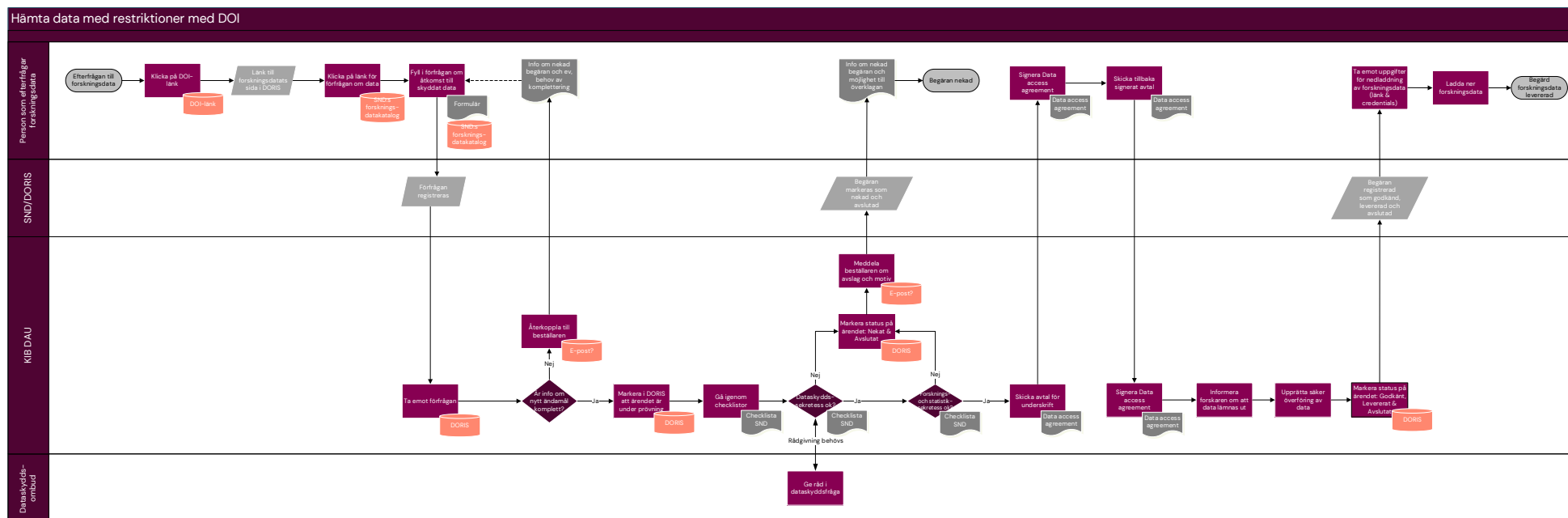
Figur 8–10 illustrerar olika processer för tillgång till forskningsdata. För hämtning av öppna forskningsdata begär användaren tillgång genom länk i katalogposten. Forskningsdata hämtas från KI:s S3-server via ett gränssnitt för åtkomst till lagringsytan.

För hämtning av forskningsdata med restriktioner (Figur 9) fyller användaren i en förfrågan. Förfrågan bedöms enligt principer i Offentlighets- och sekretesslagstiftningen och Dataskyddsförordningen. Bedömning omfattar kontroll av etikillstånd. Om data utlämnas enligt principerna för utlämnande av offentlig handling tecknas inget avtal. Om tillgängliggörandet handläggs som en datadelning regleras delningen av ett Data Access Agreement mellan KI och mottagaren av data.

Hämta öppet tillgänglig data med DOI



Figur 8, Hämta öppet tillgängliga data med DOI. Zooma in i PDF-filen för att läsa figurens detaljer.

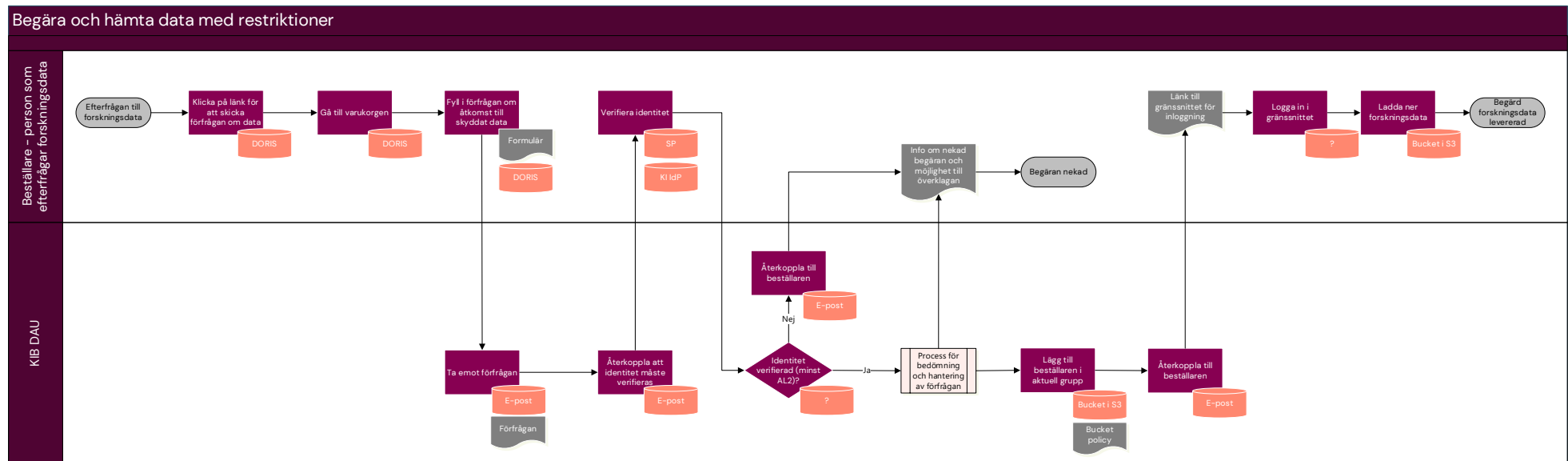


Figur 9, Hämta data med restriktioner med DOI. Zooma in i PDF-filen för att läsa figurens detaljer.

Observera att Figur 9 genererades inom projektet innan beslut om handläggningsordning för ärenden inkomna till KI via DORIS fattats. Efter beslutet renodlas processen till att antingen ske som ett utlämnande av allmän handling eller som delning av data reglerad av ett Data Access Agreement.

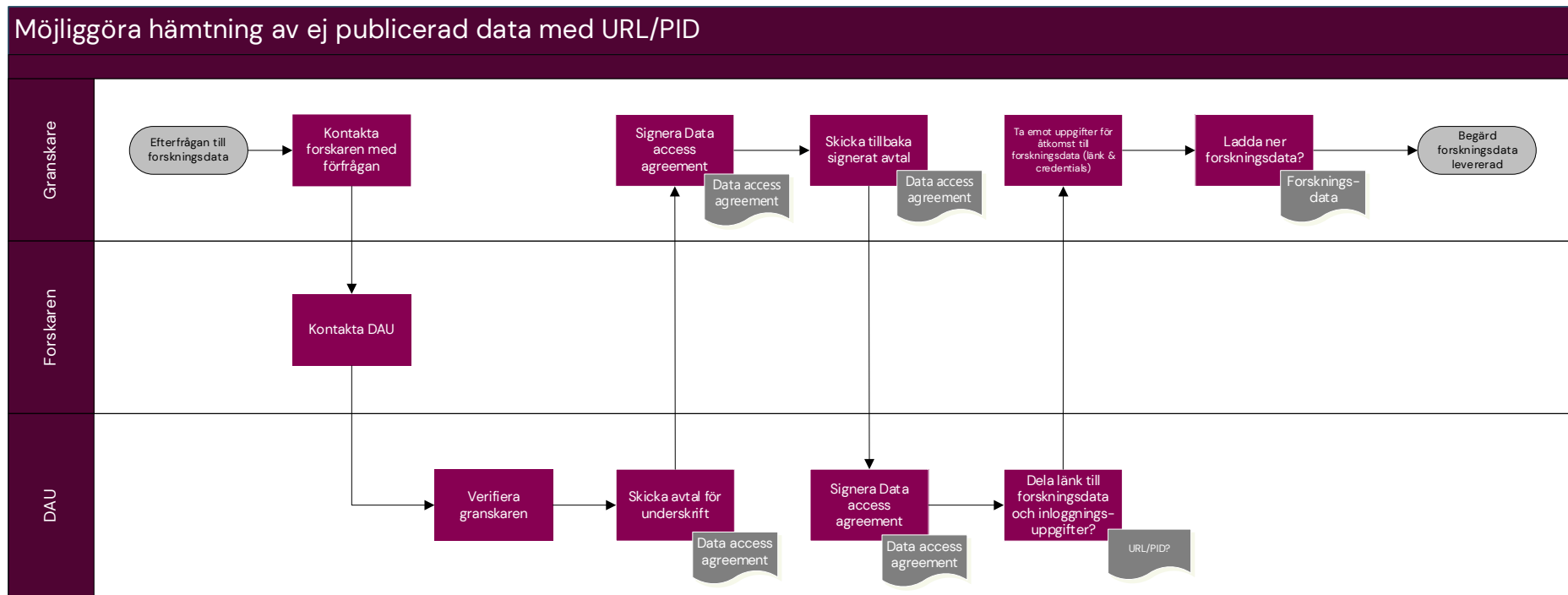
Utöver processen som Figur 9 illustrerar innehåller flödet en verifiering och autentisering av mottagarens identitet vilket illustreras i Figur 10. Baserat på information i användarens förfrågan om forskningsdata avgör DAU om en verifiering av användaren krävs för vidare handläggning av förfrågan. Verifieringen syftar till att fastställa mottagarens identitet för att kunna bedöma mottagarens syfte med dataåtkomst, verifiera att etiktillståndet gäller för avsedd dataåtkomst, och att ge tillgång till rätt person.

Processen för bedömning och hantering av förfrågan innefattar vidare att DAU granskar etiktillståndet och i vissa fall även etikansökan för att säkerställa att dataåteranvändningen omfattas av ett etiktillstånd.



Figur 10, Begära och hämta data med restriktioner, verifiering av frågeställarens identitet. Zooma in i PDF-filen för att läsa figurens detaljer.

Figur 11 illustrerar scenariot att en redaktör eller en granskare (peer reviewer) begär åtkomst till forskningsdata som ännu ej är publicerad via SND:s katalog. En förutsättning är att forskaren har förberett katalogposten för datapublicering i DORIS. Eftersom det inte går att begära forskningsdata via formuläret innan metadataposten publicerats i SND:s katalog krävs manuell hantering. Innan publicering leder den reserverade DOI:en inte till en landningssida. DAU bistår med att dela en URL till en förhandsgranskning av katalogposten och tillhörande data. Ifall att granskaren vill tillgå hela datasetet och begäran rör skyddsvärda forskningsdata bedömer DAU huruvida forskningsdata kan lämnas ut samt verifierar granskarens identitet.



Figur 11, Möjliggöra hämtning av ej publicerade data med URL/PID. Zooma in i PDF-filen för att läsa figurens detaljer.

Administrativa rutiner

Lösningsförslaget innefattar förutom teknisk lösning administrativa rutiner. Rutinernas utformning är beroende av utvecklingen av den tekniska lösningen som sker Q1–Q2 2024. Följande rutiner kommer att utvecklas:

- Rutin för hantering av forskningsdata under granskning
- Rutin för publicering av data
- Rutin för handläggning av förfrågningar
- Rutin för utlämning av data

Rutinerna baseras på redan befintliga rutiner som utarbetats i RDO:s och DAU:s löpande verksamhet. Formatet är word-dokument som lagras på DAU-gruppens gemensamma Microsoft Teams-yta. En sida på KIB:s onlinemanual/stödsystem för internkommunikation fungerar som förteckning över rutinerna.

Second line

Intern checklista för arbetsgång vid granskning, publicering och utlämning av forskningsdata i Doris och S3

Denna checklista börjar då ett dataset har skickats in till eller enbart beskrivits i DORIS. Det kan också komma in frågor via Rdo-majlen som handlar om publicering i DORIS och SND-katalogen. Då kan vi informera om vilka förberedelser överlämnaren kan göra innan hen skickar in. Följande länk kan skickas till överlämnaren: <https://snd.gu.se/sv/forbered-data-tillgangliggorande> 

- När en forskare skapat och skickat in en beskrivning så kommer ett mail att gå iväg till Rdo@ki.se. I Rdo-majlen finns en regel som flyttar meddelandet till DORIS-mappen undermapp publiceringar.
- Den som bevakar Rdo-majlen tar hand om ärendet och meddelar resten av KIB DAU

Figur 12, KI RDO:s onlinemanual på KIBsvar

Ärendehantering

I RDO:s verksamhet sker ärendehantering med hjälp av en kombination av system:

- DORIS för att motta anmälningar av registrerade katalogposter, bevakat status på poster under bearbetning, och förfrågningar
- E-post för informationsutbyte med publicerande forskare och de som begär ut data

- Ett internt register för att dokumentera tilldelade DOI:er samt att koppla dessa till var forskningsdata lagras

Vid ökat publiceringsflöde uppstår behov av ett ärendehanteringssystem.

Spårbarhet i personuppgiftshantering

För att möta föreskrifter för säkerhet i informationssystem kan rutinerna behöva kompletteras med viss loggning. Lösningförslaget har inte utrett behov av och former för loggning för spårbarhet. Följande frågeställningar behöver utredas under 2024:

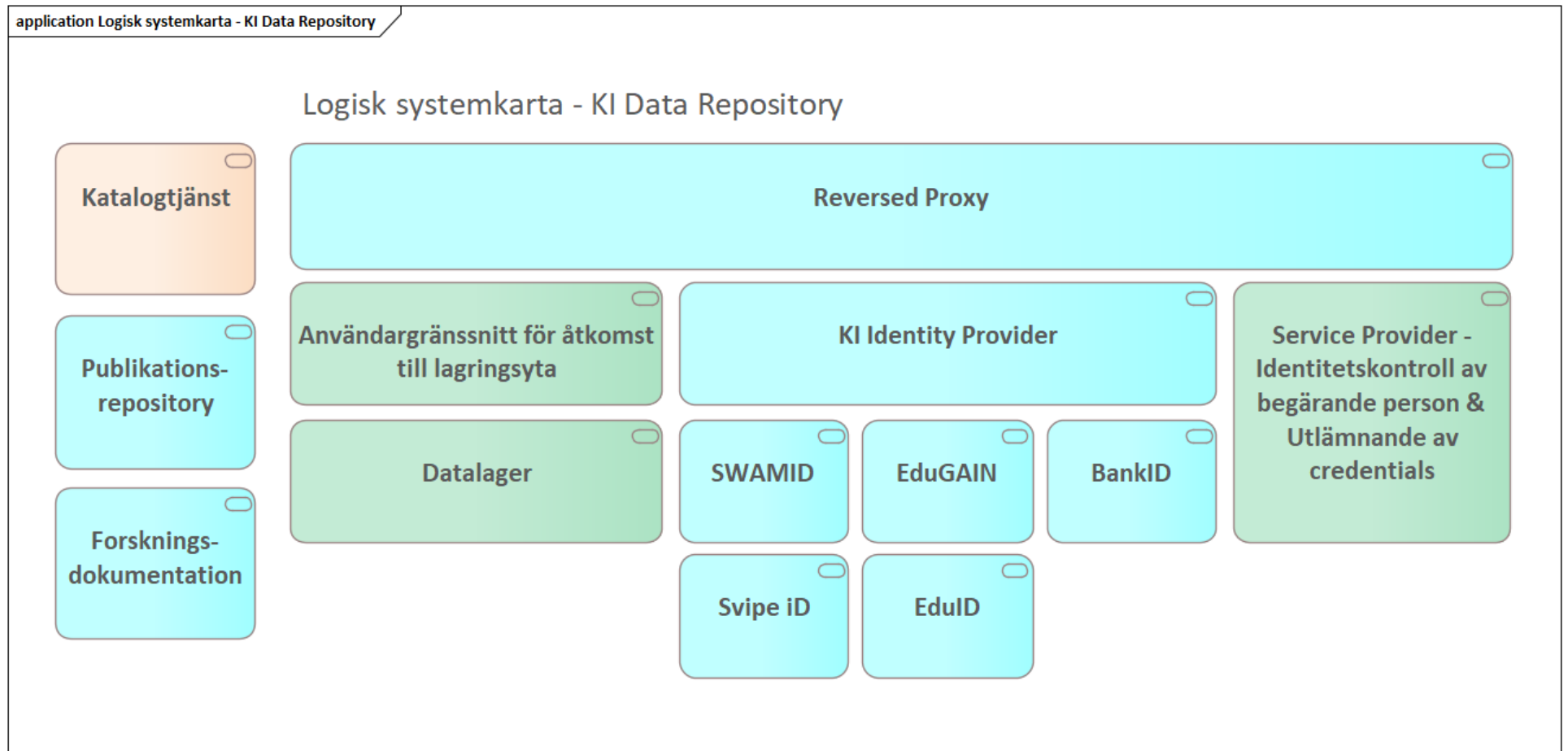
- a) Vilken information behöver loggas?
- b) Vilken information loggar befintliga system?
- c) Vilken information ska inte loggas?
- d) När ska loggad information raderas?

Möjlig teknisk lösning – Lösningsarkitektur

Följande avsnitt beskriver lösningens uppbyggnad. Inledningsvis presenterar avsnittet en logisk systemkarta som ger en överblick över den typ av komponenter som lösningen bygger på. Därefter presenteras lösningsförslaget (Figur 14).

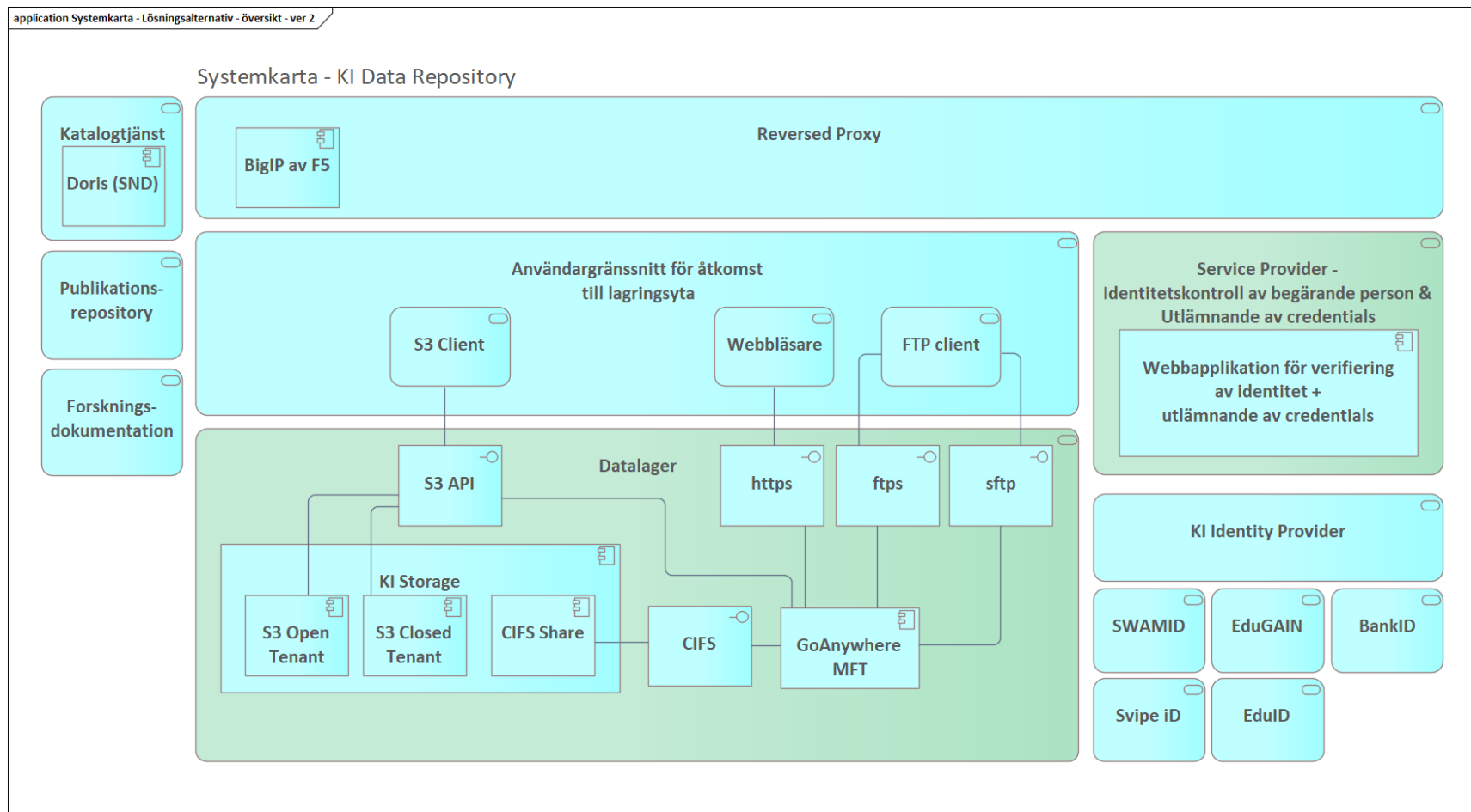
Logisk systemkarta – översikt

Figur 13, Logisk systemkarta, illustrerar typen av IT-komponenter som lösningsförslaget kräver. En kombination av komponenterna utgör lösningsförslaget.



Figur 13, Logisk systemkarta – KI Data Repository

Lösningsförslag



Figur 14, Systemkarta - Lösningalternativ - översikt

Figur 14, Systemkarta, illustrerar föreslagna IT-komponenter för lösningen. Valet av IT-komponenter är baserat på komponenter som redan används i KI:s IT-miljö. Valet av komponenter är gjort i samverkan mellan ITA och KI DAU. Komponenterna bedöms fungera för en integration med DORIS i Fas 2.

Datalager

Lösningsförslaget innebär sammanfattningsvis att datalagret utgörs av S3 och MFT. S3 kommer även fortsättningsvis användas som standardlagring, med MFT som en alternativ möjlighet.

Egenutvecklad IT-lösning för identitetskontroll

Lösningsförslaget består också av en egenutvecklad IT-lösning för identitetskontroll av begärande person. Lösningen bygger på redan befintliga lösningar för identitetskontroll genom federation respektive EduID. Identitetskontrollen krävs både för att identifiera begärande person, exempelvis att personen som begär forskningsdata med restriktioner är samma person som innehar etikillståndet eller är en person som den som innehar etikillståndet intygat har rätt att begära data för projektets räkning. Identitetskontrollen krävs också för att vid tillgängliggörandet ge begärande person åtkomst, exempelvis genom att dela ett användarnamn och lösenord.

Det finns ingen befintlig helhetslösning för den typ av identitetskontroll som KI Data Repository behöver. Alla lärosäten som ska kunna dela skyddsvärda data behöver motsvarande lösning. Utveckling av identifieringslösningen är därför relevant för SND:s konsortium och bör kommuniceras till övriga lärosäten. Det är möjligt att identifieringslösningen

framöver även kan vara användbar för andra system som hanterar forskningsinformation, till exempel KI:s e-arkiv. Projektet har inte utrett hur omfattande utveckling och förvaltning av identifieringslösningen är.

Projektet följer utvecklingen av rekommendationer för autentisering på europeisk nivå inom European Open Science Cloud:s (EOSC:s) arbetsgrupp Authentication and Authorization Infrastructure Architecture (AAI).

Användargränssnitt och användarvänlighet

Lösningsförslaget pekar inte ut något specifikt användargränssnitt för åtkomst till lagringsytan. Lagringsyta kan kommas åt på olika sätt. En lösning är att personen använder en installerad File Transfer Protocol (FTP)-klient²¹ på sin dator. Ett annat alternativ är att personen använder en webbläsare och når datalagret via GoAnywheres MFT webbgränssnitt.

Lösningsförslaget är användarvänligt för de målgrupper som har god vana att arbeta med filöverföringsklienter för filöverföringar till och från servrar. De användare som skapar och begär den typ av dataset som hanteras på detta sätt bedöms klara uppgiften med stöd av instruktioner från DAU. I kommande faser är det möjligt att initiera utveckling av förenklade användargränssnitt, dels för handläggarna på DAU för administration av lagringsytan, dels för forskarnas upp- och nedladdning till lagringsytan.

²¹ FTP är en metod för att överföra filer mellan en dator (klient) och en webbserver via Internet.

Diskussion: förslaget arbetsflöde i förhållande till aktuell lagstiftning

I föreliggande avsnitt diskuteras förslaget arbetsflöde för att beskriva och tillgängliggöra forskningsdata i förhållande till aktuell lagstiftning.

Diskussionen återger en workshop med inbjudna experter på juridik från KI och SND.

Diskussionen tar upp fyra punkter värda att uppmärksamma vid vidare utveckling av projektet KI DR: identifiering av användare, ansvaret för bedömning av om forskningsdata innehåller personuppgifter, ansvaret hos den som lämnar ut en handling, samt granskning av etikillstånd för återanvändning av data.

Identifiering av användare

Begär någon ut en allmän handling får myndigheten som huvudregel inte efterforska vem denne är eller vilket syfte personen har med sin begäran. Myndigheten får dock ställa frågor om det krävs för att kunna pröva om sekretess föreligger. Sådana frågor kan i fallet med forskningsdata vara: I vilket syfte ska data granskas eller återanvändas? Finns ett etikillstånd för återanvändningen av forskningsdatan? Det är därför viktigt att identifiering av användare inte sker förrän tillräcklig information om användarens begäran inhämtats för att kunna bedöma om det finns grund för att identifiera användaren.

Förslag:

- Förlägg verifiering av användaren efter "Process för bedömning och hantering av förfrågan", precis innan tillgång till forskningsdata (se flödet i Figur 10).

Ansvaret att bedöma om forskningsdata innehåller personuppgifter

Vid publicering och tillgängliggörande av forskningsdata via DORIS och SND:s forskningsdatakatalog sker en bedömning av forskningsdatats innehåll. Vid registrering av metadata i DORIS gör forskaren en initial bedömning av om

forskningsdata innehåller personuppgifter. Handläggare vid DAU kontrollerar forskarens bedömning. Forskaren anses ha det slutgiltiga ansvaret för att uppgifterna om forskningsdatats innehåll är korrekt angivna.

Förslag:

- Forskningsdata bör vara informationssäkerhetsklassad innan forskningsdata överförs till S3 och metadata publiceras i SND:s forskningsdatakatalog.
- Informationssäkerhetsklassning bör anges i datafilernas administrativa metadata.
- Forskare bör ges stöd att informationssäkerhetsklassa forskningsdata innan tillgängliggörande via KI Data Repository. Hänvisning till befintligt stöd för informationssäkerhetsklassning vid KI:s funktion för IT- och informationssäkerhet ska ingå i datapubliceringsrutinen.

Ansvar hos den som lämnar ut forskningsdata

Vid tillgängliggörande av forskningsdata har den tjänsteman som fått i uppdrag att ta hand om forskningsdata ansvaret att bedöma om forskningsdata kan lämnas ut enligt Offentlighets- och Sekretesslagen och Dataskyddsförordningen. Ansvaret är personligt.²² I praktiken ligger ansvaret på den eller de handläggare som delar ut rättigheter till begärande person att tillgå ett dataset.²³

Förslag:

- En noggrann granskning av forskningsdatas innehåll och av det underlag som den som begär åtkomst till forskningsdata skickar in, vid behov i samråd med forskaren eller prefekt som företrädare för forskarens institution, ska ske av handläggaren innan utlämnade.

²² "Om någon som är skyldig att hemlighålla en uppgift eller allmän handling som omfattas av sekretess lämnar ut den kan det vara brottsligt. I brottsbalken finns den centrala straffbestämmelsen om brott mot tystnadsplikt (20 kap. 3 §). Vid uppsåtligt brott mot tystnadsplikt är straffet böter eller fängelse i högst ett år och vid oaktsamt brott böter. Den som av oaktsamhet begår en ringa gärning ska inte dömas för brott." (Regeringskansliet, 2019, s. 28)

²³ "I första hand görs prövningen av om en handling kan lämnas ut av den tjänsteman som har fått i uppdrag att ta hand om handlingen, till exempel en registrator eller en handläggare i ett ärende. I tveksamma fall ska han eller hon hänskjuta saken till myndigheten, om det inte fördröjer prövningen." (Regeringskansliet, 2019, s. 21)

- Handläggaren behöver ha tillräcklig kompetens för bedömning av forskningsdatas innehåll i förhållande till Offentlighets- och Sekretesslagen och Dataskyddsförordningen.
- Vid sekretessbedömningen och granskningen av etikansökan och -tillstånd ska handläggaren ha tillgång till juridisk rådgivning.
- Granskningen av förfrågningsunderlaget och forskningsdatans innehåll i förhållande till Offentlighets- och Sekretesslagen och Dataskyddsförordningen ska dokumenteras.

Etikttillstånd för återanvändning av forskningsdata

Vid återanvändning av forskningsdata innehållande personuppgifter krävs etikttillstånd. KI har undersökningsplikt. Det innebär att handläggaren som ska lämna ut forskningsdata behöver granska att det inte finns begränsningar att ta hänsyn till i ursprungsetikttillståndet samt att etikttillståndet för återanvändningen omfattar forskning på den data som begärs ut.

Förslag:

- Figur 9, Hämta data med restriktioner med PID-DOI, bör kompletteras med begäran av etikttillstånd innan ett Data Access Agreement skickas ut.
- Handläggaren behöver ha tillräcklig kompetens för bedömning av etikttillståndets giltighet för begärda data.
- Bedömningen av etikttillståndets giltighet för begärd data ska dokumenteras.

Diskussion: föreslaget arbetsflöde och säkerhetsaspekter

I föreliggande avsnitt diskuteras föreslaget arbetsflöde för att beskriva och tillgängliggöra forskningsdata i förhållande till säkerhetsaspekter.

Diskussionen återger en workshop med KI:s experter på IT- och informationssäkerhet.

Diskussionen tar upp tre punkter värda att uppmärksamma vid vidare utveckling av projektet KI DR: att upprätthålla informationssäkerhet när forskningsdata flyttas, autentisering av användare samt tillgängliggörande av forskningsdata för externa användare.

Genomgående informationssäkerhet

Informationen som ska hanteras i lösningsförslaget ska informationsklassas. Informationen ska i första hand informations klassas med hjälp av KI:s KRTS-modell (Konfidentialitet, Riktighet, Tillgänglighet, Spårbarhet). Eftersom data avses delas utanför KI bör projektet också utreda om det finns någon lämplig internationell standard för informationsklassning. Hela flödet, från forskarens överföring av fil till lagringsytan till utdelning av forskningsdata till mottagare ska hålla en tillräcklig nivå av säkerhet för hanteringen av skyddsvärda forskningsdata.

När lösningsförslaget beslutas och implementeras ska samtliga ingående system ses över. Särskilt bör DAU-enhetens hantering av filer ses över då detta steg i processen inte tydliggörs i lösningsförslaget. I dagsläget hämtar DAU-enheten filer från MFT, sparar filer på den egna hårddisken eller, vid stora filer, på en extern hårddisk. Detta för att kunna öppna och granska att filernas innehåll överensstämmer med information angiven i metadataposten. Därefter överför DAU-enheten filerna till S3.

Förslag:

- Eftersom data avses delas utanför KI bör projektet utreda om det finns någon lämplig internationell standard för informationsklassning.

- I implementerad lösning bör ett säkert sätt för handläggaren att öppna och granska filerna ingå.

Verifiering och autentisering av användare

Användare som begär tillgång till forskningsdata behöver kunna autentiseras. Lösningen bygger delvis på federation, och delvis på Sunets Edu-ID. Edu-ID passar de personer som inte kommer in via federationen. De kan lägga upp en identitet via Sunet som sedan eleveras via Swipe-ID med stöd i 140 länder.

SND utvecklar inte identifiering och verifiering av användare. Det är lärosätets uppgift att ha kontroll över vilka som tillgår data. Därför behöver KI lokalt utveckla verifierings- och autentiseringslösningen. Detsamma gäller för alla lärosäten som ska kunna ge tillgång till skyddsvärda forskningsdata. Det finns risker knutna till att utveckla och förvalta system lokalt. Det skulle finnas stora fördelar om SND-konsortiet nationellt gick samman för att utveckla lösningen för lokal implementering.

Identifieringen och verifieringen av användare behöver spåras, liksom utdelning och, efter bestämd tid, återtagande av rättigheter att hämta forskningsdata. För detta behövs någon typ av ärendehanteringssystem som stödjer struktur för behörigheter. Manuella system medför risker att uppgifter inte läggs in eller att ärenden inte följs upp enligt rutiner.

Förslag:

- KI behöver utveckla identifieringslösningen.
- Identifieringen och verifieringen av användare, liksom utdelning och återtagande av rättigheter att tillgå forskningsdata behöver vara spårbart.
- För spårbarhet i hanteringen behövs ett ärendehanteringssystem som stödjer struktur för behörigheter.

Tillgängliggörande av forskningsdata för extern användare

För tillgängliggörande av forskningsdata till personer som förfrågar forskningsdata delas rättigheter att tillgå forskningsdata som ligger på KI:s S3-server. Forskningsdata ligger i olika tenants och buckets, beroende på om forskningsdata är öppen eller om tillgången ska begränsas, till exempel för att data innehåller personuppgifter. Till öppna forskningsdata länkas direkt från katalogposten i SND:s forskningsdatakatalog. Till forskningsdata med begränsad tillgång får användare tillgång via inloggningsuppgifter. Tillgängliggörandet ska ske med en verifierad användare och hämtning ska vara möjligt under en begränsad tid. Det interna arbetsflödets och tillgängliggörandets spårbarhet är viktig.

Det finns en hypotetisk risk med länkar som leder in i KI:s IT-miljö. Om länken går att manipulera så kan det vara en väg in i KI:s miljöer. Kontroll över behörighetstilldelningen är avgörande för att stävja denna risk.

Förslag:

- Tillgängliggörandet ska ske med en verifierad användare och hämtning ska vara möjligt under en begränsad tid.
- Det interna arbetsflödets och tillgängliggörandets spårbarhet är viktig.

Referenser

Europaparlamentet och Europeiska unionens råd. (2019).

Europaparlamentets och Rådets direktiv (EU) 2019/1024 av den 20 juni 2019 om öppna data och vidareutnyttjande av information från den offentliga sektorn. <https://eur-lex.europa.eu/legal-content/SV/TXT/PDF/?uri=CELEX:32019L1024&from=EN>

KI (2024) *Policy för öppen tillgång till forskningsdata*

Projektbeställning KI Data Repository – Fas 1 (2022-12-07) (internt dokument, KI)

Projektplan KI Data Repository – Fas 1 (2023-11-01) (internt dokument, KI)

Regeringskansliet (2019) *Offentlighetsprincipen och sekretess – Kortfattat om lagstiftningen.* Stockholm: Regeringskansliet.
<https://www.regeringen.se/informationsmaterial/2019/06/offentlighet-principen-och-sekretess--kortfattat-om-lagstiftningen/>

SCB (2011) *Standard för svensk indelning av forskningsämnen.*
<https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-indelning-av-forskningsamnen/>

SND (2023a) *Flaggskepp* <https://snd.gu.se/sv/flaggskepp>

SND (2023b) *Interimslösning för egen lagring utan integration med DORIS (Version 3).* Zenodo. <https://doi.org/10.5281/zenodo.7785310>

SND (2022a). *Integrationen mellan lärosätets egen lagring och SND (DORIS) (Version 9).* Zenodo. <https://doi.org/10.5281/zenodo.6346298>

SND. (2022b). *Vägledning för utlämnande av forskningsdata med personuppgifter (Version 1).* Zenodo.
<https://doi.org/10.5281/zenodo.6352658>

SND (2021a). *Beskrivning av DAU-funktionen (Version 5).* Zenodo.
<https://doi.org/10.5281/zenodo.6346126>

SND (2021b) Fil: Arbetsflöden 2.png—DAU-handboken.
https://dhub.snd.gu.se/wiki/Fil:Arbetsfl%C3%B6den_2.png

Sunet (2020) *Identity Assurance Level 2 Profile—Sunet Wiki*.

<https://wiki.sunet.se/display/SWAMID/Identity+Assurance+Level+2+Profile?preview=/78218828/83494424/SWAMID%20Identity%20Assurance%20Level%202%20Profile%20v2.0%20FINAL.pdf>

Vetenskapsrådet. (2022) *Lär dig mer om öppen tillgång till forskningsdata*.

Vetenskapsrådet. Stockholm: Vetenskapsrådet.

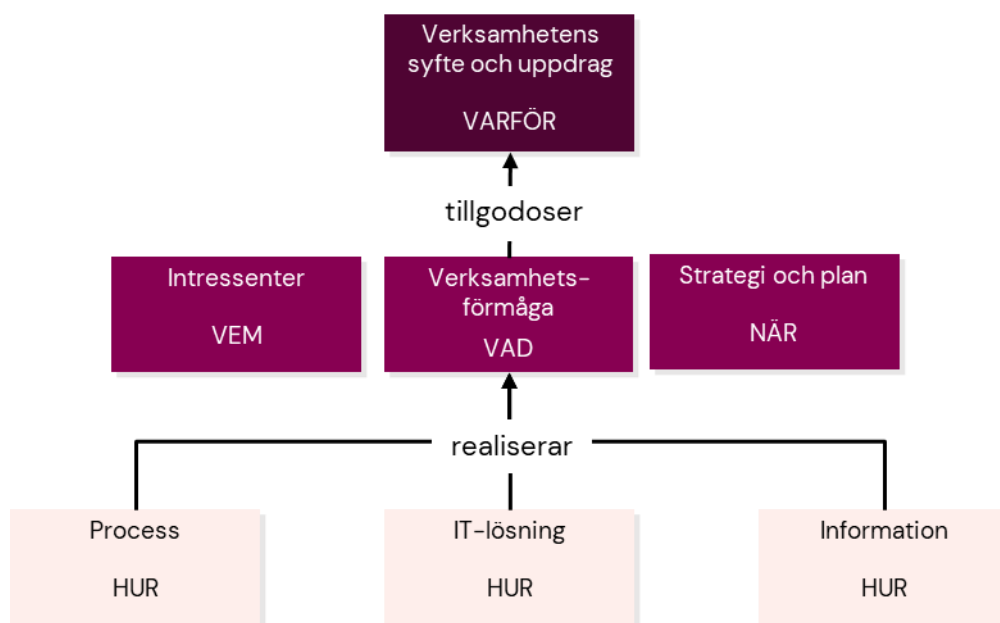
Bilaga 1, Grundprinciper verksamhets- och lösningsarkitektur

För att hitta bästa möjliga IT-stöd inleds arbetet med att bygga en arkitektur med att identifiera och beskriva verksamhetens syfte och behov. Det vill säga **varför** verksamheten finns och hur verksamheten ger nytta till KI.

För att en verksamhet ska kunna uppfylla sitt syfte krävs olika typer av **verksamhetsförmågor**. Det vill säga **vad** en verksamhet gör och **kan**.

Hur verksamhetsförmågorna förverkligas, med olika typer av resurser och lösningar, beskrivs med arkitekturella modeller som delas in i tre olika kategorier: Processer, IT-lösningar och Information.

Utöver modeller som beskriver *varför*, *vad* och *hur* kompletterar en *intressentkarta* (som beskriver **vem** som på olika sätt påverkas av lösningarna) de arkitekturella modellerna.



Figur 15, Grundprinciper för verksamhets- och lösningsarkitektur

Arkitekturens fyra nivåer

Arkitekturen beskrivs i fyra nivåer:

- Verksamhetsarkitektur: Beskriver exempelvis processer och information, utan att ha fokus på IT-lösningar
- Lösningsarkitektur: Beskriver IT-lösningar på en grov nivå (till exempel vilka IT-system som används i en process)
- Mjukvaruarkitektur: Beskriver IT-lösningar mer detaljerat (till exempel hur specifika IT-system implementerats)
- Infrastrukturarkitektur: Beskriver nödvändig infrastruktur (till exempel nätverk och serverplattformar)

Respektive nivå beskrivs utifrån tre olika perspektiv

- IT-lösningar & Komponenter – beskriver ingående system och komponenter
- Processer & Roller – beskriver arbetssätt, vilka roller som deltar och vem som gör vad
- Begrepp & Information – beskriver information som används

	IT-lösningar & Komponenter	Processer & Roller	Begrepp & Information
VERKSAMHETS- ARKITEKTUR	Logiska system/applikationer Strukturer	Processmodell, Processbeskrivning, Aktivitetsdiagram Roller och ansvar	Begreppsmodell, Begrepp, Informationsmodell, Informationsobjekt
LÖSNINGS- ARKITEKTUR	Logiska system, Referensarkitektur, Användningsfallsöversikt, IT-tjänster, Systemkarta, Applikationsdiagram	Grova jobbflöden	Logisk datamodell
MJUKVARU- ARKITEKTUR	Klassdiagram, Web services, Komponenter inom en applikation	Specat dialogflöde, Prototyp, Sekvensdiagram, Batchjob-specar	Fysisk datamodell
INFRASTRUKTUR- ARKITEKTUR	Plattformar, Fysiska servrar, Virtuella servrar, Nätverk, Databashanterare	Kommunikationsspecar inkl. protokoll och metoder	Databas, Tabell

Figur 16, Arkitekturs fyra nivåer, detaljerad beskrivning