



Karolinska
Institutet

KI Data Repository – Fas 1, delrapport 1

SND:s Flaggskepp 2023–2024

Flaggskeppsrapport: <https://zenodo.org/records/10990448>

Frågor: rdo@ki.se





Forskare på Karolinska Institutet behöver publicera och lagra forskningsdata på ett säkert och effektivt sätt.

Många av KI:s forskare arbetar med data med känsliga personuppgifter.



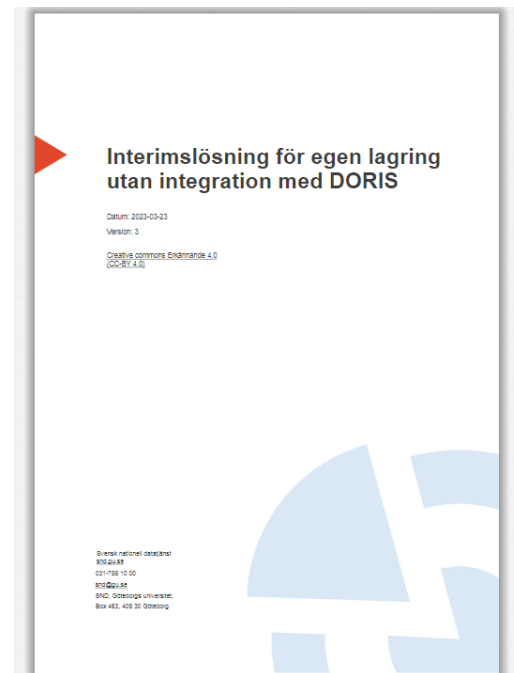
Lagring, publicering och tillgängliggörande av data regleras av lagstiftning och riktlinjer.



KI DR skapar en robust teknisk infrastruktur för centralt stöd till KI:s forskare att öppet tillgängliggöra metadata och forskningsdata.

Projektet KI Data Repository (KI DR)

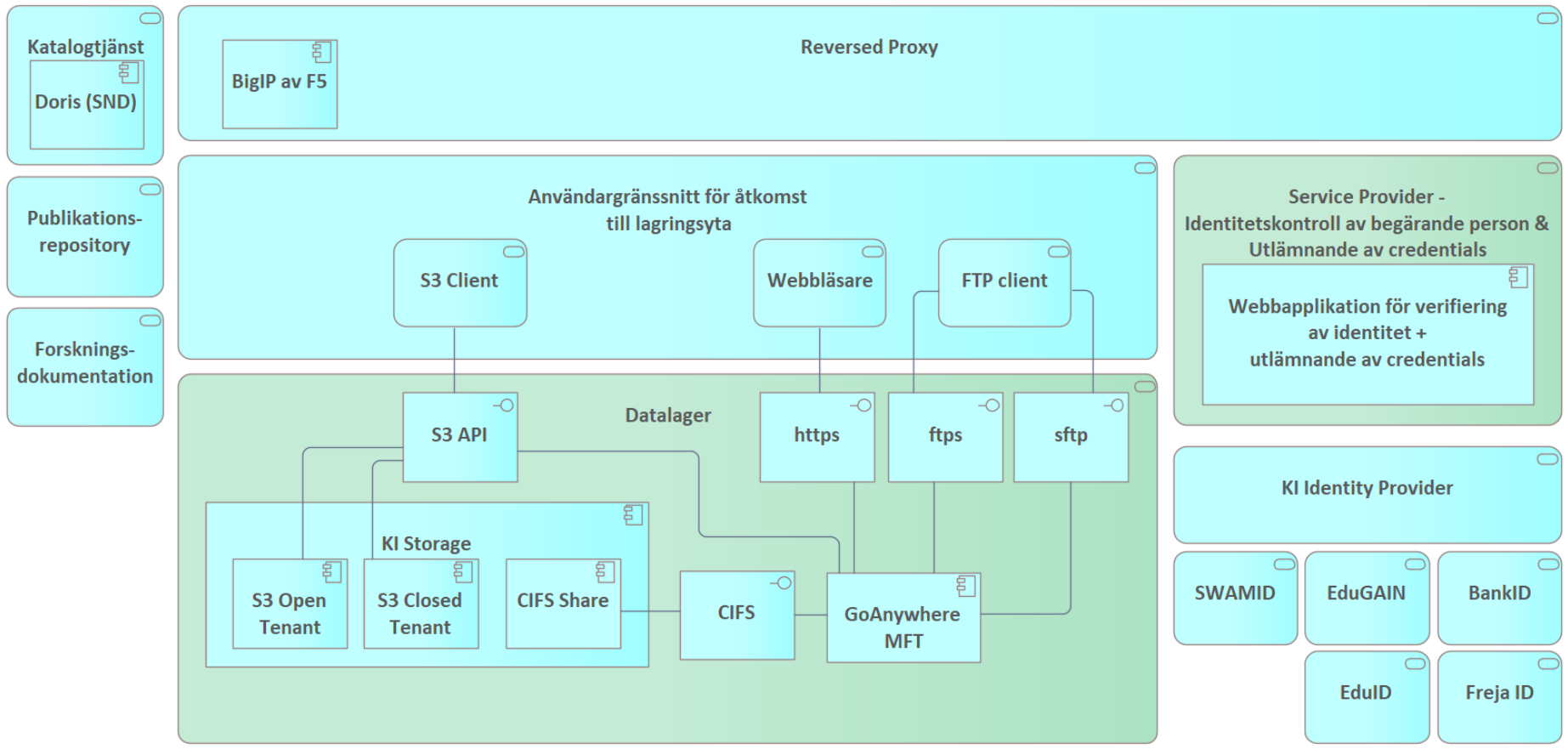
#	Leverans
L1	Förslag/modell till ett komplett arbetsflöde som omfattar både en möjlig teknisk lösning och de nödvändiga administrativa rutinerna.
L2	Redovisa eventuella konflikter mellan målsättningen att skapa ett välfungerande arbetsflöde för att beskriva och tillgängliggöra forskningsdata med skyddsvärda 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) handlar om förslag till fungerande arbetsflöden, både administrativt och tekniskt.
L4	Fungerande teknisk lösning.
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.



Projektets bidrag till verksamheten i Fas 1

- Anvisning till gränssnitt för att ladda upp forskningsdata
 - Yta för öppna data
 - 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 säkerhetsklassade ytor
 - Öppna data
 - Skyddsvärda data
- Möjlighet att generera länkar till öppna forskningsdata
- Förmågan att verifiera och autentisera identitet vid utlämning av skyddsvärda data
- Metod för utlämnande av skyddsvärda data (rutin + teknisk lösning)
- Spårbarhet i utdelning och återtagande av rättigheter

Systemkarta - KI Data Repository



Utmaning – Autentisering

- MFA (MultiFaktors Autentisering) stöds inte av S3 i nuläget.
- Lösningar KI undersöker i fortsatt arbete.
 - SciShare och NextCloud

Utmaning – Data Integritet

- Bästa alternativet: Använd Baggit lokalt för att generera checksums.
- Önskemål: Verifiera checksummor i varje steg. Utmaning – beräkna checksummor för filer på S3.
- Etags genereras per automatik under uppladdningen till S3 men de är inte identiska med MD5 checksums.

Utmaning – Användarvänlighet vid överföring av stora filer

Stängda datasets

- Användning av open source-klienter med tydliga instruktioner
- Begränsat stöd för kommandoradsverktyg
- Behövs ett webbaserat gränssnitt?

Öppna datasets

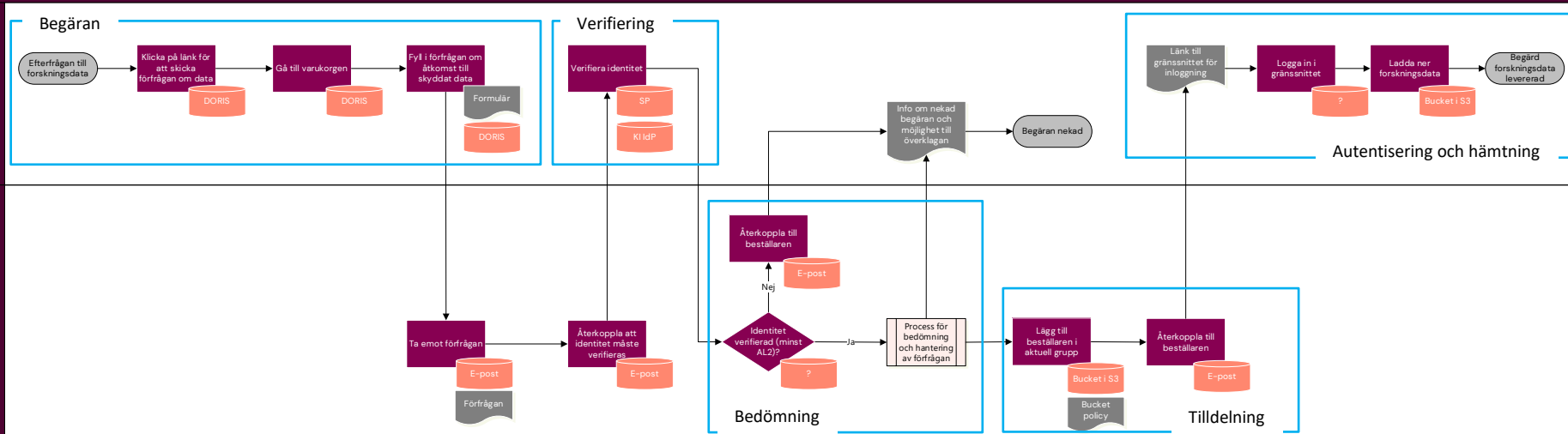
- Möjlighet att länka till en static webbsida på S3 med index av filerna
- Zip filerna.

Autentisering

Begära och hämta data med restriktioner

Beställare - person som efterfrågar forskningsdata

KIB DAU



På gång

- Formellt beslut om handläggningsordning (Kommittén för forskning)
- Fortsätta utveckla teknisk lösning för säker hantering, lagring och utdelning av data
- Utveckla teknisk lösning för autentisering och verifiering
- Utarbeta rutiner och system för spårbarhet

- Core Trust Seal-certifiering
- Ta fram plan för långsiktig finansiering
- Information till forskare
- Planera för integration med DORIS (Fas 2)



**Karolinska
Institutet**

Diskussion: förslaget arbetsflöde i förhållande till aktuell lagstiftning

- Identifiering av användare
 - Myndigheten får inte efterforska vem som begär ut en allmän handling eller vilket syfte personen har med sin begäran.
 - Myndigheten får ställa frågor om det krävs för att kunna pröva om sekretess föreligger.
 - Identifiering av användare bör inte ske förrän tillräcklig information om användarens begäran inhämtats.
- Bedömning av forskningsdata innehåll
 - Informationssäkerhetsklassning innan överföring till lagring
 - Ange informationssäkerhetsklassning i administrativa metadata
 - Stöd till forskare för informationssäkerhetsklassning

Diskussion: förslaget arbetsflöde i förhållande till aktuell lagstiftning, forts.

- Ansvaret för utlämnande av forskningsdata ligger på tjänstemannen
 - Handläggaren måste bedöma om forskningsdata kan lämnas ut enligt Offentlighets- och Sekretesslagen och Dataskyddsförordningen
 - Handläggaren behöver ha tillräcklig kompetens för bedömning av forskningsdatas innehåll
 - Vid sekretessbedömningen och granskningen av etikansökan och -tillstånd ska handläggaren ha tillgång till juridisk rådgivning
 - Granskningen av förfrågningsunderlaget och forskningsdatans innehåll ska dokumenteras
- Etikillstånd krävs för återanvändning av forskningsdata innehållande personuppgifter
 - KI har undersökningsplikt
 - Handläggaren behöver granska att det inte finns begränsningar att ta hänsyn till i det ursprungliga etikillståndet

Diskussion: föreslaget arbetsflöde och säkerhetsaspekter

- Genomgående informationssäkerhet
 - 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
 - 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.

Diskussion: föreslaget arbetsflöde och säkerhetsaspekter, forts.

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