



Krav och rekommendationer för data och metadata i SND:s forskningsdatakatalog

Datum: 2021-09-13

Version: 3

[Creative Commons Erkännande 4.0](#)

(CC-BY 4.0)

Svensk nationell datatjänst


snd.gu.se

031-786 10 00

snd@gu.se

SND, Göteborgs universitet,

Box 463, 405 30 Göteborg



Inledning

Svensk nationell datatjänst (SND) är en nationell infrastruktur för att beskriva, dela och återanvända forskningsdata. SND är ett konsortium bestående av nio lärosäten.¹

SND:s forskningsdatakatalog har som huvudsakligt syfte att göra forskningsdata sökbara, åtkomliga och återanvändbara. Data och metadata som delas via forskningsdatakatalogen ska i så hög utsträckning som möjligt uppfylla FAIR-principerna, så som principerna tillämpas inom det forskningsområde för vilket datamaterialet är framställt. De krav och rekommendationer som presenteras i det här dokumentet utgår från detta syfte.

Dokumentet riktar sig till två målgrupper:

1. organisationer i SND-nätverket som har egna lagringslösningar kopplade till SND:s system
2. SND CARE² (repositoriedelen för SND:s verksamhet).

Att känna till SND:s system är en förutsättning för att väl kunna tillgodose sig informationen i det här dokumentet.

Organisationer som har egna lagringslösningar för forskningsdata kopplade till SND:s system är som forskningshuvudmän ansvariga för de data och metadata som delas via SND:s forskningsdatakatalog. Kontroller ska säkerställa att de data och metadata som delas i forskningsdatakatalogen uppfyller de krav som anges i det här dokumentet.

Data som tillhör en organisation som *inte* har egen lagring kopplad till SND:s system hanteras via SND CARE. För data som hanteras av SND CARE är SND som organisation ansvarig för hantering och publicering av data och metadata, detta gäller även om en DAU-funktion från forskningshuvudmannen medverkar vid kontroll av data och metadata. För SND CARE ska även rekommendationerna i det här dokumentet uppfyllas då de är en del i SND:s certifiering som Trusted Digital Repository enligt CoreTrustSeal.

Dokumentets syfte och upplägg

I detta dokument finns krav och rekommendationer för hur data och metadata som ska göras sökbara i SND:s forskningsdatakatalog ska kontrolleras och eventuellt berikas. Dokumentet täcker kontroll av data och metadata samt juridiska förutsättningar för att dela data.

Organisationer i SND-nätverket, som har egen lagring kopplad till SND:s system, kan välja att på organisationsnivå ge sina forskare möjlighet att göra en s.k. *snabbregistrering av data*³ i SND:s

¹ Chalmers, Göteborgs universitet, Karolinska institutet, Kungliga tekniska högskolan, Lunds universitet, Stockholms universitet, Sveriges lantbruksuniversitet, Umeå universitet och Uppsala universitet.

² SND CARE är SND:s CoreTrustSeal-certifierade (<https://www.coretrustseal.org/>) repositorieverksamhet.

³ I SND:s ansökan till Vetenskapsrådet (2021) omnämns den här funktionen som *entry level data*.

system. Data som beskrivs på den här nivån omfattas av särskilda krav, vilka presenteras sist i respektive avsnitt nedan.

Som komplement till det här dokumentet finns DAU-handboken⁴ som bland annat innehåller praktiska beskrivningar för hantering och kontroll av data och metadata, samt förslag på arbetsflöden och rutiner som stöd för det arbete som ska göras hos forskningshuvudmannen. DAU-handboken finns tillgänglig online och utvecklas löpande av både DAU- och SND-medarbetare.

Kontroll av metadata och dokumentation

De krav som listas nedan är framtagna med syfte att uppfylla SND:s mål om att så långt som möjligt uppfylla FAIR-principerna, samt för att uppfylla krav från centrala aktörer och samarbetspartners (som exempelvis DataCite⁵).

Kraven på metadata gäller både databeskrivningar där data nås via SND:s forskningsdatakatalog (de är direkt nerladdningsbara eller kan beställas via en förfrågan) och för databeskrivningar där data enbart är beskrivna i SND:s katalog med en hänvisning till en annan portal/aktör för åtkomst till data. Kraven avser däremot *inte* katalogposter vars metadata automatiskt speglas/höstas till SND-katalogen från andra portaler, t.ex. Språkbanken och SITES. För skördade metadata bör kraven på metadata ses som ett rikt-
märke och kan behöva justeras från fall till fall.

Kraven omfattar såväl övergripande metadata som metadata på fil- eller variabelnivå.

Krav för metadata och dokumentation

- Obligatoriska metadatafält i formuläret i DORIS ska vara ifyllda och beskrivna så utförligt som möjligt för de aktuella data. Obligatoriska metadatafält ska också vara beskrivna på både svenska och engelska, med undantag för fältet *Titel*.⁶
- I den mån det finns artiklar och/eller andra publikationer relaterade till data ska metadata innehålla länkar/referenser till dessa.
- I den mån externa länkar finns ska dessa, om möjligt, vara av typen persistenta identifierare.
- Tillhörande dokumentation⁷ som är relevant för att förstå och återanvända de aktuella data ska delas tillsammans med data och metadata. Information om materialets framtagande, bearbetning och sammanhang bör finnas med.

⁴ <https://dhd.snd.gu.se>

⁵ <https://datacite.org/>

⁶ DORIS är SND:s dokumentationsverktyg för forskningsdata. Vilka fält som är obligatoriska avgörs av vald metadataprofil i formuläret. Kontrollerade vokabulärer och nyckelordslistor översätts automatiskt i DORIS, fritextfält kräver manuell översättning. Översättningarna behöver inte vara exakta, men likvärdig information ska finnas på båda språken. Vid längre texter kan en hänvisning till katalogposten för det andra språket vara aktuellt, men en inledning bör finnas på båda språken.

⁷ För att ett dataset ska kunna återanvändas för ny forskning eller kunna reproduceras är tillhörande dokumentation viktig. Med tillhörande dokumentation avses här information i löpande text som inte kan hanteras enbart via standardiserade metadataelement.

Att definiera vad som är tillräcklig dokumentation för återanvändning är beroende av bland annat forskningsområde, typ av data och det specifika forskningsprojektet. En vägledning finns i DAU-handboken, men det krävs en egen bedömning i varje enskilt

I de fall databeskrivningen omfattas av data som delas med avseende på en specifik artikel (eller annan publikation) kan kravet om tillhörande dokumentation frångås, villkorat att den aktuella artikeln/publikationen:

- är öppen och fritt tillgänglig
- är länkad till i databeskrivningen med en persistent identifierare (PID)
- innehåller all information som krävs för att data ska gå att förstå och återanvända, så att resultat ska kunna återskapas oberoende av andra resurser.⁸

Rekommendationer för metadata och dokumentation

- Ej obligatoriska men för forskningsområdet viktiga metadatafält bör vara beskrivna i så stor utsträckning som möjligt.
- Metadata och tillhörande dokumentation bör, i så hög grad som möjligt, vara förståbara för forskare från andra ämnesområden. För högt specialiserade material kan det vara rimligt att acceptera en snävare målgrupp (t.ex. genetiska data).
- Metadata för dataset som publiceras i samband med en specifik artikel bör vara beskrivna så att resultaten i artikeln kan replikeras.
- Om en specifik artikel behövs för att återanvända data (se sista punkten i Krav för kontroll av data) bör artikeln sparas tillsammans med data för att säkerställa framtida åtkomst.⁹
- Om data delas via någon annan portal/aktör än SND, d.v.s. enbart metadata delas via SND:s forskningsdatakatalog, bör relevant dokumentation finnas hos den andra portalen/aktören eller tillsammans med datafilerna. Oavsett om sådan dokumentation finns eller inte måste SND:s PID-policy uppfyllas.

Krav på metadata för snabbregistrerade data

Profilen för snabbregistrerade data utgörs av ett mindre antal metadataelement och uppfyller inte de generella kraven för SND:s forskningsdatakatalog. Följande krav gäller för metadata för snabbregistrerade data:

- Obligatoriska metadatafält i formuläret i DORIS ska vara ifyllda och beskrivna så utförligt som möjligt för de aktuella data. Obligatoriska metadatafält ska också vara beskrivna på båda svenska och engelska, med undantag för fältet *Titel*.¹⁰

fall om data har den dokumentation som är nödvändig. Exempel på dokumentation är variabellista, frågeformulär och teknisk rapport.

⁸ Observera att det i de allra flesta fall behövs omfattande dokumentation för att data ska kunna återanvändas och att en enskild artikel inte alltid är tillräckligt. Ofta behövs kompletterande dokumentation vid sidan av artikeln om t.ex. innehållet i datafilen/-erna. För en numerisk datafil kan det t.ex. vara en readme-fil, variabellista eller kodbok som beskriver de variabler och koder som finns i datafilen.

⁹ Artikeln behöver inte delas tillsammans med data så länge det finns en fungerande länk till artikeln. Men artikeln bör finnas sparad för framtiden om länken skulle brytas.

¹⁰ Se fotnot 6.

- I den mån det finns artiklar och/eller andra publikationer ska metadata innehålla länkar/referenser till dessa.
- I den mån externa länkar finns ska dessa, om möjligt, vara av typen persistenta identifierare.

SND:s krav på metadata för snabbregistrerade data säkerställer att DataCites krav på metadata uppfylls så att data kan tilldelas en DOI (Digital Object Identifier), samt att det finns tillräcklig information om datas innehåll så att filerna kan behandlas korrekt ur ett säkerhetsperspektiv. Möjligheterna att återanvända och att uppfylla FAIR-principerna är lägre för forskningsdata med dessa begränsade beskrivningar. Målsättningen bör därför vara att data ska beskrivas mer utförligt i ett kommande steg.

Kontroll av data

Nedan redogörs för de krav och rekommendationer som gäller för kontroll av data. Kraven är framtagna för att säkerställa datafilernas kvalitet och innehåll, så att data kan förstås och återanvändas av andra.

Krav på kontroll av data gäller enbart databeskrivningar där data nås via SND:s forskningsdatakatalog, d.v.s. de är direkt nerladdningsbara eller kan beställas via en förfrågan.

Krav för kontroll av data

Forskningsdata som delas via SND:s forskningsdatakatalog ska ha genomgått följande kontroller:

- Det material som ska delas och/eller beskrivas ska vara forskningsdata. Med forskningsdata avses här digitalt material som kan ligga till grund för en vetenskaplig analys oavsett forskningsområde.
- Filer ska kontrolleras för virus.
- Filer ska finnas i lämpliga format för återanvändning och långtidsbevarande. Om data och/eller filformat är väldigt specifika för forskningsområdet och kräver särskild programvara för att öppnas och användas, kan det i undantagsfall vara tillräckligt att forskaren intygar filernas riktighet och vilka format som är lämpligast.
- Om särskild programvara krävs ska den medfölja data eller beskrivas på ett sådant sätt att den går att hitta. Programvarans namn och version ska framgå samt ytterligare information som är relevant för återanvändning.
- Har kod eller script använts för att bearbeta eller tolka data ska det framgå och kod/script ska bifogas datafilerna.¹¹ Information om eventuell programmeringsmiljö, version, och annat som kan vara relevant för återanvändning ska framgå.
- Filpaketet är komplett – den innehåller all data och dokumentation som är nödvändig för korrekt återanvändning.

¹¹ Det ska finnas information om hur man kör scripten. Scripten ska ej innehålla absoluta sökvägar (använd relativa sökvägar).

- Innehållet i datafilerna ska kontrolleras med avseende på personuppgifter eller annan skyddsvärd information. Mer om kontroll av personuppgifter i avsnittet om *Juridiska förutsättningar för att dela data* nedan.

Rekommendationer för kontroll av data

- Filer (inklusive eventuella mappar) bör vara namngivna på ett konsekvent och begripligt sätt. Detta är särskilt viktigt för omfattande material med många filer. Filnamn, innehåll och relationer mellan filer kan behöva förklaras i ett kompletterande textdokument.
- Filer bör vara rensade från ovidkommande information (t.ex. formateringar och exempelvariabler som inte används eller som saknar betydelse för forskningsresultatet).

Krav på kontroll av data för snabbregistrerade data

För kontroll av data omfattas snabbregistrerade data av samma krav som övriga databeskrivningar:

- Filer ska kontrolleras för virus.
- Filer ska finnas i lämpliga format för återanvändning och långtidsbevarande. Om data och/eller filformat är väldigt specifika för forskningsområdet och kräver särskild programvara för att öppnas och användas, kan det i undantagsfall vara tillräckligt att forskaren intygar filernas riktighet och vilka format som är lämpligast.
- Innehållet i datafilerna ska kontrolleras med avseende på personuppgifter eller annan skyddsvärd information. Mer om detta i avsnittet om *Juridiska förutsättningar* nedan.

Juridiska förutsättningar för att dela data

Utöver de ovan angivna kontrollerna behöver varje forskningshuvudman se till att det finns rutiner för att säkerställa att hanteringen av data uppfyller de krav som ställs i Dataskyddsförordningen (GDPR) och annan relevant lagstiftning.

Nedan följer ett klargörande av hur ansvaret för forskningsdata och användningen av SND:s forskningsdatakatalog är fördelat och vilka kontroller som är nödvändiga för att säkerställa att katalogen används på ett korrekt sätt.

I SND:s forskningsdatakatalog finns två tillgänglighetsnivåer:

1. Direkt nedladdningsbara forskningsdata.
2. Ej direkt nedladdningsbara data med restriktioner, där utlämningen måste prövas av forskningshuvudmannen innan utlämnande.

Det åligger varje forskningshuvudman att bedöma innehållet i alla forskningsdata, så att rätt tillgänglighetsnivå tillämpas. I normalfallet görs detta av forskaren och DAU vilket beskrivs nedan, men respektive forskningshuvudman kan också ha beslutat om andra rutiner. Det är dock viktigt att

säkerställa att samtliga verksamheter har rutiner och kunskap på plats så att SND:s forskningsdatakatalog inte används i strid med gällande dataskyddslagstiftning. Det är också av stor vikt att forskningsdatakatalogen inte används så att forskningsdata hanteras och sprids på ett sätt som utgör ett brott enligt nationell och internationell lagstiftning eller i strid med avtal. Med brott i det här sammanhanget avses bl.a. förtal, röjande av sekretessbelagda uppgifter, upphovsrättsintrång, spridning av barnpornografiskt material eller andra brottsbelagda handlingar. Med ”i strid med avtal” avses till exempel spridning av forskningsdata som reglerats i avtal med exempelvis en extern tredje part där rättighetsanspråk eller konfidentialitet regleras särskilt.

För att minimera risken för att SND:s katalog används på ett lagstridigt sätt är det därför nödvändigt att rutinerna för att göra data tillgängliga innehåller vissa kontroller och att forskare och DAU har tillräckliga kunskaper för att genomföra dessa kontroller. Dessa rutiner består dels av att den enskilde forskaren intygar att val av tillgänglighetsnivå för data är förenlig med relevant lagstiftning samt inte utgör brott, dels av att DAU-funktionen utför en egen kontroll av forskningsdata för att säkerställa korrekt tillgänglighetsnivå samt att data i övrigt delas enligt gällande lagstiftning. Nedan beskrivs översiktligt hur dessa kontroller går till i DORIS.

Observera att processen för själva utlämnandet av data med restriktion, exempelvis sekretessprövning, ej beskrivs. En prövning måste göras av forskningshuvudmannen innan utlämnande sker.

Kontroll ett - forskaren

Oavsett tillgänglighetsnivå måste forskaren fylla i en försäkran genom kryssrutor i DORIS att forskningsdata inte innehåller uppgifter vars spridning eller tillgängliggörande utgör brott och inte innehåller uppgifter vars spridning eller tillgängliggörande strider mot avtal.

Forskaren som väljer att göra data tillgängliga i SND:s forskningsdatakatalog utan restriktioner behöver även kryssa i en försäkran om att data inte innehåller personuppgifter eller andra sekretessbelagda uppgifter, som normalt inte får delas öppet.

Kontroll två - DAU

I den granskning DAU utför behöver det även ingå att kontrollera att forskaren kryssat i ovan nämnda försäkran enligt *kontroll ett* på ett korrekt sätt. Därefter behöver DAU-funktionen även göra en egen bedömning av att vald tillgänglighetsnivå av forskningsdata är förenlig med relevant lagstiftning.

Detta innebär att göra en kontroll utifrån samma kriterier som forskaren, det vill säga att vald tillgänglighetsnivå är förenlig med dataskyddslagstiftningen samt att publiceringen inte utgör brott.

Förutsättningar för lagring av data

Data och tillhörande filer som nås via SND:s forskningsdatakatalog (de är direkt nerladdningsbara eller kan beställas via en förfrågan) måste finnas lagrade på ett kontrollerat och långsiktigt sätt. Det finns i dagsläget två alternativ för lagring av data:

1. Data lagras hos SND CARE. SND är ansvariga för att hanteringen av data och övriga filer sker på ett korrekt sätt.
2. Data lagras på yta/-or som kontrolleras av forskningshuvudmannen. Organisationer som ingår i SND-nätverket har möjlighet att koppla egen lagring till SND:s system. Detta skapar förutsättningar för att dela data med skyddsvärd information, exempelvis personuppgifter, via SND:s forskningsdatakatalog, eftersom data stannar hos forskningshuvudmannen.

Det finns inget krav från SND om detaljerna kring forskningshuvudmannens lagring, ansvaret för lagring ligger på respektive organisation. Varje organisation är själv ansvarig för att utveckla rutiner för hur forskningsdata lagras.