

Kommentarer til rapporten som helhet

Takk for anledningen til å komme med innspill til utkastet til rapport.

Samlet sett utgjør rapporten en oppsummering av kunnskapsgrunnlaget som i stor grad var til stede før utredningen startet, og gir noen gode case-eksempler på tilgjengeliggjøring av data. Sammenstillingen kan ha verdi som introduksjon til feltet, for forskere, forskningsledere og rådgivere i generelle forskningsstøttefunksjoner, som ikke har jobbet inngående med problemstillingene. Gjennomgangen kunne vært mer aktuell dersom man hadde plassert åpne data i en større sammenheng av åpen forskning.

Når det er sagt, bør rapporten underlegges en gjennomgående redigering med gjennomgående kvalitetssikring, og revidering av endel spesifikke beskrivelser og resonnementer. Dette må til for at rapporten skal få et presisjonsnivå som gjør at den kan fungere som grunnlagsdokument og introduksjon til status på feltet. Utover den grunnleggende informasjonsverdien rapporten kan få etter noen nødvendige korrigeringer, mener NTNU at rapporten i begrenset grad kan bidra til utvikling, siden rapporten ikke tydelig nok identifiserer behovet for nye eller forbedrede infrastrukturer og tjenester.

Rapportens anbefalinger er relevante, men åpenbare, og er ikke tilstrekkelig konkrete til at de kan stimulere til utvikling. Grunnlaget for anbefalingene må gjøres tydelig (er det utredningsgruppens eller referansegruppens egen kompetanse og erfaringer, eller spørreundersøkelsen). Anbefalingene bør konkretiseres og gis tydelig adresse med plassering av ansvar for oppfølging. Anbefalinger om praksis og kultur hos forskere og hva institusjonene bør oppfordre til er selvfølgelig viktige aspekter, selv om dette er "gammelt nytt". Det bør knyttes anbefalinger til utvikling eller samarbeid om konkrete verktøy og infrastruktur.

NTNU er enig i at kompetanse i alle ledd, at kultur for deling er viktig, men dette er ikke sentralt i mandatet. For ordens skyld vil vi bemerke at forslaget på omfang av datarøktene er ganske urealistisk, for eksempel vil dette for NTNUs vedkommende innebære langt over 100 slike stillinger.

En annen generell kommentar er at noen begreper brukes uten å defineres, eksempelvis, "sensitive data" og "deling av forskningsdata". Dette er begrep som kan brukes og oppfattes på ulike vis, avhengig av kontekst. For bedre lesevennlighet foreslår vi å korte ned kapittel 5 og ta det med i introduksjonen. Kapittel 4 og 6 kan slås sammen, slik at det først beskrives en mangel/et behov, og deretter følger konkrete anbefalinger om hvordan dette behovet kan dekkes (på overordnet nivå).

NTNU mener at det i større grad bør legges vekt på de konkrete behovene som er fremkommet, for eksempel behovet for DMP-verktøy med bedre funksjonalitet, sentralt juridisk kompetansesenter eller lignende, behovet for elektronisk labjournal, og behovet for bedre løsninger for lagring (samhandling) og arkivering av persondata og annen type data som krever beskyttelse og tilgangskontroll. Dette vil gjøre det mulig å komme med konkrete forslag og anbefalinger om hvilke utfordringer som kan og bør løses i fellesskap.

NTNU støtter sektorsamarbeid om tjenester, verktøy og infrastruktur som legger til rette for FAIR forskningsdata. Vi mener sektoren vil være tjent med standarder og felles kriterier, og at institusjonene gjør ting på samme måte. Dette er ikke det samme som at alle nye løsninger skal utvikles og organiseres i en sentralisert eller nasjonal struktur, selv om felles systemer i noen tilfeller er hensiktsmessig.

I tillegg til NTNUs ovennevnte innspill, har Research Data@NTNU listet endel spesifikke kommentarer og justeringer til de enkelte kapitlene, se nedenfor. De følgende kommentarene er forslag til revidering av rapporten slik at en anvendelse i introduksjonsøyemed for forskere og forskningsstøttegeneralister, får tilstrekkelig presisjonsnivå. NTNU bistår gjerne med ytterligere spesifikasjoner til den endelige rapporten.

Research Data @NTNUs innspill til rapport fra utredningen *Felles infrastruktur og tjenester for FAIR forskningsdata*.

Research Data @NTNU er NTNUs sentrale støttetjeneste for forskningsdatahåndtering. De følgende kommentarene er våre konkrete forslag til detaljert revidering av de enkelte kapitlene i rapporten.

Kommentarer til kapittel 2 – Innledning

- 2.3; inkluder gjerne en beskrivelse av referansegruppene.
- Side 7, første avsnitt: Det bør komme tydeligere frem at metadata bør opprettes for alle typer data, og om det er data som ikke kan deles åpent, bør likevel metadata og info om datasettet deles åpent.
- Side 14: Figur om “Increasingly FAIR” bør presiseres med tanke på forskjell på åpen og FAIR.
- Side 15: Definisjon av lagring vs arkivering bør presiseres og nyanseres, for eksempel er ikke lagring et begrep som brukes om kun “aktive data”, det beskriver generelt lagring på et medium, og tilsvarende, arkiverte data er ikke nødvendigvis alltid “kalde”.
Klassifisering: Flere aspekter (i tillegg til personopplysninger) er relevante med tanke på konfidensialitet.
- Side 16: Definisjon av datakurerering bør klargjøres.
- Generelt vil det være nyttig for lesere med flere direkte lenker til ressursene som beskrives, eksempler er styringsdokumentet til RDA (s 7), opendsi.org (s 8) og alle dokumentene i dokumentstudiet/referanselisten

Kommentarer til kapittel 3 – Hvilke tjenester og verktøy brukes i dag?

Kapittel 3 er, slik vi ser det, ikke grundig nok for å avdekke konkrete behov for verktøy, og hva som kan være nyttig av videre utvikling av tjenester og funksjonalitet. Kapittelet presenterer i stor grad allerede kjent kunnskap om forskeres praksis for datahåndtering. Nedenfor er listet noen formuleringer som bør justeres i endelig versjon av rapporten.

3.2.1. Søke etter eksisterende data

"...viser vi til oversikten over arkiver som brukes av forskerne i dag" – her trengs en henvisning.

3.2.2. Lage datahåndteringsplan

"Av de som hadde erfaring hadde flest brukt mal fra egen institusjon og NSDs DMP var det mest brukte verktøyet». Dette er uklart og ulogisk, da NSDs verktøy foreløpig ikke tillater institusjonelle maler. Er det institusjonsmaler eller NSDs verktøy som er mest brukt blant forskere? Videre bør eksemplene på datahåndteringsplanverktøy inneholde *verktøy*, og fortrinnsvis de som er relevante for forskere i Norge.

3.3. Håndtere forskningsdata i pågående prosjekter

Første avsnitt om hovedfokus er ikke konsistent med delkapittelet for øvrig. Eksemplene fremstår som et tilfeldig, og snevert, utvalg av tjenester. I og med "hovedfokus" er på lagring av data, savner vi en vurdering av hvordan lagringstjenestene fungerer for forskningsdata, med en sammenligning av fordeler og ulemper. Når det gjelder EBRAIN, er dette et veldig godt eksempel på en infrastruktur for forskning, mer enn et eksempel på en tjeneste for aktiv lagring.

Metadata: Flere verktøy for metadata bør nevnes i tillegg til elektronisk labjournal, sammen med beskrivelse av hvordan verktøyene bør integreres i arbeidsprosesser for å legge til rette for FAIR data.

3.4. Arkivere og publisere forskningsdata

Listen over eksempler inneholder tjenester som ikke hovedsaklig er arkiv eller som i alle fall primært har andre hovedtjenester knyttet til lagring, analyse etc (NeIS, HUNT Cloud).

Det hadde vært verdifullt å tydeliggjøre læringspunkter og nytte ved de ulike casene. Eksempelvis case om UNGDATA (side 24), hva har blitt lært av dette caset som er relevant for andre og for ev utvikling av (felles) tjenester og verktøy eller beste praksis og prinsipper?

3.5. Tilgang til kompetanse og bistand

Open Science Toolbox: Presisere at dette er et verktøy for å utvikle forskningsstøtte ved fagbibliotek, knyttet til Åpen vitenskap og forskningsdatahåndtering.

Kommentarer til kapittel 4 – Hva mangler?

Kapittel 4 bør utdypes på noen punkter med større presisjon. Nedenfor er noen innspill til konkretisering.

4.1. Verktøy og tjenester

- Kulepunktene inneholder spesifikasjoner som de fleste arkiv allerede tilbyr. Rapporten bør nevne hvilke verktøy/tjenester er det snakk om, for eksempel når det gjelder manglende API, om det er spesifikke behov i ulike fagfelt, og hva som mangler i konkrete tjenester.
- "Fleksibel tilgangskontroll" bør beskrives mer inngående eller med annet begrep, «gradert» kan være en mulighet, da det ligger i selve prinsippet at tilgang kan være ulik for ulike brukere. Informasjonssikkerhet er generelt overflatisk behandlet i hele dokumentet.
- TSD er strengt tatt ikke et arkiv. Vi er enig i at det er behov for tjenester for samhandling om fortrolige data, og arkivering av disse, utover NSD.
- "Flere av institusjonene arbeider for tiden med å etablere egne infrastrukturer for deling av forskningsdata" bør konkretiseres mht om deling gjelder arkivering/publisering eller annen deling. Konkrete eksempler på prosjekt/tiltak savnes.
- «Behov for forutsigbare kostnader»: Det er et klart behov for forutsigbare forretningsmodeller der kostnader er kjent for prosjektleder og prosjekteier i planleggingsfasen av prosjektet. Infrastruktur

som er nødvendig for å overholde lovkrav, for eksempel arkivering av forskningsdata iht. Helseforskningslovens bestemmelser, bør være gratis for forsker/forskningsprosjektet.

4.2. Kunnskapsbehov og kompetanse

- Kunnskap og kompetanse er relevant, men ikke sentralt i mandatet.
- Side 36 (nederst): *Dersom forskerne pålegges å publisere sine data på denne måten vil disse dataene også omfattes av åpne data-direktivet og offentlighetsloven. Hva betyr dette?*

4.3. Håndtering av data med behov for beskyttelse

- Schrems II-dommen gjør det pr. nå vanskelig (men ikke umulig) å dele persondata med land utenfor EU. Det gjelder ikke bare USA men de fleste land utenfor EU/EØS.
- Det er behov for støtte til forskere for å avklare enkelte gråsoner mellom hva som regnes som ikke-sensitive, fortrolige og strengt fortrolige data, og for hva som regnes som anonyme eller pseudonyme data. Videre er det behov for støtte/infrastruktur som sørger for aidentifisering og anonymisering av identifiserbare sensitive data. Anonymisering etter prosjektslutt er et virkemiddel som kan bidra til bedre tilgjengeliggjøring og dermed økt gjenbruk av data.

4.6. Oppsummering av dagens utfordringsbilde

Oppsummeringen innfører nye moment som ikke står i sammenheng til resten av kapittelet, for eksempel "små, enkle tjenester for analyse", "nasjonale maler".

Kommentarer til kapittel 5 – Diskusjon

Som nevnt i det overordnede innspillet fra NTNU, bør dette kapittelet forkortes og tas inn som introduksjon. Videre har vi følgende kommentarer:

- Det er ikke tilstrekkelig tydelig hvem "vi" er.
- «Langtidsbevaring» bør defineres.
- NTNU er enig i at det er heterogenitet, og dette bør kunne ut i at det blir identifisert konkrete behov.
- Side 40 (nederst): NTNU imøteser leveransen med oversikt over tjenester som er tilgjengelig for norske forskningsinstitusjoner, da denne oversikten ikke er inkludert i rapporten.

Kommentarer til kapittel 6 – Anbefalinger

NTNU mener dette er den viktigste delen av rapporten. Anbefalingene bør konkretiseres og gis tydelig adresse med plassering av ansvar for oppfølging. Det bør også knyttes anbefalinger til utvikling eller samarbeid om konkrete verktøy og infrastruktur. Grunnlaget for anbefalingene må gjøres tydelig (er det utredningsgruppens eller referansegruppens egen kompetanse og erfaringer, eller spørreundersøkelsen).

6.1. Forskningens bruk av verktøy og tjenester

Mange av anbefalingene generelle, og er kjent stoff, som gir grunnlag for inntrykket av at rapporten kan (med nødvendige korreksjoner) bli en grei introduksjon til feltet. Noen av anbefalingene har som tidligere

nevnt uklar adresse, og er ikke tilstrekkelig konkrete til at de kan stimulere til utvikling. Nedenfor følger noen forslag til presiseringer.

Verktøy for datahåndteringsplan: «Institusjonene bør anbefale ... verktøy .. som har funksjonalitet (APIer) ..». Her bør nevnes hvilke verktøy det er snakk om.

Metadata: «Institusjonene bør anbefale bruk av tjenester som har god funksjonalitet for å hjelpe forskerne til å legge på metadata...» Her bør nevnes hvilke verktøy som finnes, eller som det er behov for.

Håndtering av aktive data: Skillet mellom åpne data, persondata og sensitive data er ikke presist. Det er ikke korrekt at sensitive data *skal* lagres i godkjente nasjonale eller internasjonale tjenester. Mange institusjoner har gode systemer for lagring av data som krever beskyttelse. Utfordringen ligger i at disse ikke åpner for samarbeid på tvers av institusjonene (NICE1, SAFE etc.) eller at de ikke gir tilstrekkelige muligheter for analyse og behandling av data. Det er uheldig at rapporten anbefaler en konkret leverandør/tjeneste (TSD fra UiO). Det må presiseres hva som menes med "godkjente" løsninger (godkjent av hvem, for hvem, etter hvilke kriterier).

Arkivering av data: Avsnittene om persondata må korrigeres.

- «I en del tilfeller kan persondata (men ikke sensitive data) arkiveres hos NSD, men dette forutsetter samtykke.» Tillegg: Også persondata som anses som sensitive («særlige kategorier») kan arkiveres hos NSD dersom det foreligger samtykke.
- «Det mangler gode nasjonale generiske infrastrukturer for arkivering og langtidsbevaring av persondata generelt og sensitive data spesielt». Korrigering: Det som mangler, er i stor grad gode løsninger for begrenset tilgang og tilhørende samtykker som spesifiserer hvordan tilgangen administreres.
- «Dersom det ikke er mulig å publisere dataene åpent, vurder om det er mulig å publisere metadata, slik at personer med legitim grunn til å få tilgang, kan oppdage dataene og søke om slik tilgang». Presisering: Dette forutsetter at det finnes gode løsninger for tilgangsstyring.

6.2. Forbedring og utvikling av verktøy og tjenester

Verktøy for datahåndteringsplan: Her savner vi presisering av hvordan leverandører bør utvikle tjenestene.

Generelle anbefalinger for videreutvikling av tjenester og verktøy for håndtering av forskningsdata:

Nedenfor gir vi noen forslag til større presisjon i rapporten:

- «Funksjonalitet for å utstyre dataene med FAIR metadata» bør spesifiseres.
- «Leverandører må sikre at datasett som deles påføres ... (PID/DOI)». Eksisterende nasjonale tjenester (NSD, NIRD, BIRD, DataverseNO) har dette allerede.
- «Det bør vurderes å etablere (evt utvide eksisterende) tjenester nasjonalt for arkivering og langtidsbevaring av persondata/sensitive data». Adressaten for dette må klargjøres.
- «Det kan eventuelt vurderes større grad av samarbeid mellom institusjonene om (anskaffelse av) elektronisk labjournal.» Vi er enig i at dette er viktig, og adressaten må klargjøres.
- Faktaboksen om NVA hører ikke hjemme i kapittelet med anbefalinger.

6.3. Kompetanse

- Ref tidligere kommentar; kompetanse i alle ledd er viktig, men dette er ikke sentralt i mandatet.

- «Nasjonale aktører» må presiseres. «Regulatoriske sandkasser» er ikke nevnt tidligere i rapporten, og det er uklart hva som menes.

6.4. Kultur for deling

Vi er enig i at kultur for deling er viktig, og det bør inkluderes i introduksjonen sammen med viktigheten av insentiver for deling.