UiO : Universitetet i Oslo

Innspillsrunde om FAIR forskningdata

Jeg svarer

på vegne av en institusjon/organisasjon

Kontaktinformasjon

Navn på institusjon/organisasjon

Korbinian Michael Bösl

Navn på lavere enhet/forskergruppe (hvis relevant)

ELIXIR Norge

Navn på personen som svarer

korbinian.bosl@uib.no

E-post til kontaktperson

korbinian.bosl@uib.no

Hvis du svarer på vegne av en enhet/forskergruppe eller som enkeltperson: Hvilke(t) fagområde jobber du innen?

Det er mulig å krysse av for flere.

Matematikk/naturvitenskap

Medisin og helsefag

Landbruks- og fiskerifag

Om du har oppgitt "Annet", vær vennlig å spesifisere

Ikke besvart

Publisere og presentere innspill

I utgangspunktet har vi ønske om å publisere alle innspill på egen nettside. For enkeltpersoner er det mulig å be om at dette ikke skjer. Ønsker du at ditt innspill skal publiseres?

Ja

Presentasjon på innspillsseminar

Er du interessert i å presentere innspillet i et kort (digitalt) innlegg på innspillsseminaret 19. januar? (3-5 min avhengig av hvor mange som melder seg.)

Ja

Nedenfor er en liste med noen av de spørsmålene vi gjerne vil ha innspill på. Dette er ment som ideer til dere som leser. Dere står fritt til å kommentere både dette og andre sider ved rapporten.

Mangler det viktige elementer i beskrivelsen av dagens tjenester?

Mangler det viktige elementer i beskrivelsen av manglene i dagens tjenester og behovene for å forbedre disse?

Er det andre problemstillinger eller anbefalinger som burde vært løftet?

Er det problemstillinger eller anbefalinger som burde vært fjernet?

Har du/dere forslag til case som burde være med for å belyse problemstillinger knyttet til

forskningsdata på en god måte?

Har du/dere forslag til case på miljøer som har gjort litt i retning FAIR, som ikke er super-eksperter, men som kan være gode forbilder for de som ønsker å komme i gang?

Dere kan gi innspill enten ved å fylle ut i feltene nedenfor, ved å legge ved en fil med kommentarer i (f.eks. Pdf-fil med kommentarer eller Word-fil med «spor endring») eller som en kombinasjon.
Har du spørsmål til utfyllingen, send e-post til kristin.braenden@agendakaupang.no

Legg eventuelt ved fil her:

Har du behov for å legge ved flere filer, kan disse sendes på e-post til kristin.braenden@agendakaupang.no

Ikke besvart

Kommentarer til rapporten som helhet (ikke spesifikt til ett kapittel)

While we fully agree on the general importance of the work performed in this committee, we would like to point out several weaknesses we have identified on the work process as well as on the results.

We think the timing of calls for contribution from researchers and institutions was suboptimal both regarding the initial surveys and the current consultation. Deadlines during or directly after summer and Christmas breaks are not suited to increase the participation rate and we think there might be no good representation from different fields in the responses.
Furthermore we think that the survey for researchers is poorly designed and leads to misleading results. This includes mixing of guiding resources and DMP tools (e.g. the RDMkit is not a DMP tool, but a knowledge resource), mislabeling of metadata management tools (e.g. CEDAR) as DMP tools and labeling of storage/processing platforms (e.g. TSD) as archiving solutions.

We believe that both in the surveys and the report the topic of metadata management is not addressed and missing to large extents.

We have identified several inaccuracies and potential mistakes on topics and tools within our service domain, which we would like to address below for each chapter.

Kommentarer til kapittel 2 – Innledning

p.6 We think (molecular) Life Sciences could be also listed as one of the disciplines with a long data sharing tradition with big societal impacts and implications.

p.9 Nettsider under openscience.no: One major challenge of this platform is the lack of open contribution processes e.g. through reviewed pull requests. In the current form, openscience.no will not scale and will be difficult to maintain. The content itself is also delivered without a standard license at the moment.

Retningslinjer/kriterier som tjenester knyttet til forskningsdata bør oppfylle: We think that it will very

difficult for the institutions to provide update guidance beyond generic recommendations on a domain level. This should be included in the considerations.

p.10 We would like to criticize the design and timing of the survey. This includes mixing of guiding resources and DMP tools (e.g. the RDMkit is not a DMP tool, but a knowledge resource), mislabeling of metadata management tools (e.g. CEDAR) as DMP tools and labeling of storage/processing platforms (e.g. TSD) as archiving solutions.

p 12 “DEFINISJONER”
We think it is potentially useful to define the concept of “linked data”, since this is the main pillar for sharing machine-readable (meta)data on the web.

p.13
We are missing explicit mention of "globally unique and persistent identifier" which is an explicit requirement from F1 and F3

Deskriptive metadata
We think the considerations around metadata are missing requirements towards standardization and machine readability, together with community standards. These are also represented in the FAIR principles R1.1 and R1.3 The concept of "controlled vocabularies / ontologies " is missing, this is required by I2

Administrative metadata
Licenses are a requirement and not optional (R1.1)

Begrepet metadata
The concept of "provenance" should be made explicit (R1.2)

p.14
“[...] og viser at man kan definere mange grader eller trinn fra fullstendig lukkede data til helt FAIR data – og hvordan små grep kan gjøre at lukkede data mer tilgjengelige, selv om de fortsatt ikke er helt FAIR.”
This sentence conveys the impression that closed (lukkede) and FAIR are opposite concepts. This is not true, as data can be fully open and not FAIR at all.

p.15 We would like to dispute the definitions of storage and archiving. Storage during a project can be hot or cold, depending on the project requirements. Archives can also serve hot data for direct use (e.g. in the ELIXIR data deposition databases) See also: Chapter 7.4.4 Mons, B. Data Stewardship For Open Science: implementing FAIR principles. (2018).

This paragraph seems to imply that there must be a strong bias towards what is archived and what it

is not, which is not necessarily desirable (high-quality data that does not support a scientific hypothesis and it is thus not included in a publication should in principle be archived as well).

Archiving includes storage with special requirements for metadata and infrastructure. Metadata has to be gathered sufficiently before deposition, the approach to gather metadata at the point of deposition is too late from our experience. Further archives require infrastructure and effort for curation which is not mentioned here either.

p.16
The distinction between red and black should be made more clear in general. From general/institutional guidelines, it is hard to place a borderline between these two categories for researchers and support personnel from our experience.

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

p.18
Domain and datatype specific repositories are often the first place researchers start their search, but are missing from the list completely.

Re3data is an index of repositories and does not provide a direct dataset search, hence we think it is listed in the wrong context.

We are missing the search services for molecular life science datasets e.g. https://www.omicsdi.org/ and the underlying https://www.ebi.ac.uk/ebisearch/overview.ebi/about service

p.19

We think this section has to emphasize the value of a DMP for planning and budgeting a project. The hard requirements from funders and the evaluation rubric from e.g. Science Europe are not mentioned at all.

We would like to suggest to change the entry ‘DMP fra Data Stewardship Wizard’ to ‘Data Stewardship Wizard fra ELIXIR ESFRI’

We would like to mention that the RDMkit is not a DMP tool, but a knowledge resource and should not be mixed with those, it could be listed with e.g. the CESSDA Data Management Expert Guide

p.20
We strongly believe metadata collection has to start at the collection and analysis phase, but seems to be omitted here.
p.21
The data storage locations is highly influenced by the data classification, but this is not represented

in the survey and the text. Institutional guidelines could also be mentioned in this context (e.g. https://www.uio.no/english/services/it/security/lsis/storage-guide.html for UiO)

p.22 We would suggest to precise the entry ‘ELIXIR – ressurser og plattformer for livsvitenskapsdata’ to ‘ELIXIR Norway - ressurser og plattformer for livsvitenskapsdata (Norwegian e-infrastructure for Life Science, integrert med NIRD Data Storage)’

Furthermore we feel the section on metadata is completely omitting domain and technology specific standards (FAIR R1) and the listed services are more focused towards data storage and not management of metadata. This could be expanded with e.g. CEDAR, FAIRDOM SEEK and COPO for metadata management.
Also the metadata requirements at established domain and technology specific repositories are not mentioned in this section.

Good metadata requires a semantic foundation (e.g. through ontologies), which should be included in the text. These are also part of the necessary infrastructure and require maintenance.

p.25 Eksempler på fagspesifikke tjenester for arkivering og/eller publisering:

We suggest to remove the Norwegian eInfrastructure for Life Science (NeLS) as it is a storage and analysis/processing platform and should be mentioned on p. 22

We would like to suggest to include the Norwegian Node of the Federated European Genome Phenome Archive (FEGA) as an archive for sensitive human genetic and phenotypic data

p27
ELIXIR box
We believe the leading role of ELIXIR in providing and recommending services covering the data life-cycle for life sciences should be better defined. The ELIXIR core data resources (https://elixir-europe.org/platforms/data/core-data-resources) and recommended interoperability resources (https://elixir-europe.org/platforms/interoperability/rir-selection) can potentially be mentioned in this context. Moreover, it should be made clearer that initiatives of this type promote alignment at an European and international level and hinder the fragmentation that could occur e.g. from pure institutional or national initiatives.
In the ELIXIR Norway description, the helpdesk can also be mentioned, as it is very relevant to the “TILGANG TIL KOMPETANSE OG BISTAND” topic.
p.27
In the context of courses, ELIXIR’s training portal TeSS (https://tess.elixir-europe.org/) could also be mentioned.

p.28

We would be happy if “Elixirs informasjonsside for forskere med informasjon om rolle og nyttige ressurser” could be rephrased to: “RDMkit side for forskere med informasjon om rolle og nyttige ressurser”
We would like to emphasize that while the RDMkit is originating from an ELIXIR project it also includes efforts from other Life Science ESFRIs such as Euro-Bioimaging and BBMRI

Oversikt over arkiver: Could also list FAIRsharing which is a cross domain index of archives, standards and CVs/ontologies run by ELIXIR-UK and is recommended e.g. in the EU Horizon Europe Program Guide.

Kommentarer til kapittel 4 – Hva mangler?

p.30 Enkle, fleksible tjenester: We are missing resources for domain specific needs among the listed requirements

Fleksibel tilgangskontroll: We would like to mention the GA4GH passport in this context, which can be used with ELIXIR-AAI for tiered access control to datasets

Tjenester for lagring og arkivering av sensitive data: ELIXIR Norway har utviklet og lanserer i 2022 en arkiveringstjeneste for sensitive persondata for livsvitenskap, som bygger på TSD som plattform. FEGA Norway vil muliggjøre FAIR-deling med kontrollert tilgang til sensitive persondata som genetiske sekvenser og metadata om individer. Tjenesten er en del av det Europeiske Federated EGA-nettverket som lanseres i 2022.

p.32 Viktigheten av tidlig planlegging og innsats
We would like to mention in this context the guidance on consent form design in the RDMkit Human Data pages and further information from GA4GH and BBMRI on this topic.

p.33 Persondata

We experience repeatedly that there is a lack of knowledge about the nature of the data/technologies in regard to its sensitivity and that researchers should have clarified this at the start of the project, a service to evaluate this at the publication stage might be too late. A sufficient degree of familiarity with the technologies and type of data is necessary to perform such an evaluation.

Behov for mer kunnskap om metadata

We would like to stress that the solutions provided within the commercial LIMS/ELN landscape are mostly not compatible with the existing metadata standards and the focus should be rather open source solutions with support for ontologies.

We are also missing here the mentioning of metadata management platforms, in contrast to ELNs such as FAIRDOM SEEK, COPO, transSMART, CEDAR

p.37 Behov for verktøy og tjenester for FAIR forskningsdata
We would like to warn against a too high degree of fragmentation of DMP templates, also due to the difficulties of maintainability. We would rather recommend an approach using suggestions on top of one decision-tree based template which can build on each other.

Metadata
“Det er behov for tjenester der man kan publisere søkbare metadata uten å
publisere datasett og behov for verktøy som hjelper forskerne i å legge på og
vedlikeholde/ha kontroll på metadata fra oppstart til avsluttet prosjekt.” - We think this paragraph should be rephrased to represent the requirements presented earlier in the report, including information on how re-users can access the data.

We would like to add that metadata standards also have to allow reuse within the same field and have to have a sufficient amount of details.

We would like to refer to the NeLS tool assembly https://rdmkit.elixir-europe.org/nels_assembly.html as an example of a non-fragmented service for RDM

p.38
Behov for informasjon og kompetanse
“Flere forskere
kjenner imidlertid ikke til verktøy for datahåndteringsplan, lagring og arkivering av data.”

We would like to stress again existing information resources including the RDMkit to inform support personnel and researchers

We also think the ongoing activities by the carpentries should be considered in this context.

Kommentarer til kapittel 5 – Diskusjon

p.41 anbefale funksjonalitet i datahåndteringsplaner som gjør det mulig
ågjenbruke og dele informasjonen senere i prosessen

We are trying to work actively towards this goal with NeLS, DSW and FAIRDOM SEEK - however many of the DMP tools listed above are not suited for machine actionability

Kommentarer til kapittel 6 – Anbefalinger

p.44 Forskere bør bruke internasjonale metadatastandarder der det finnes.

We would like to refer also to https://fairsharing.org/ as a cross domain index for archives,

(meta)data standards, controlled vocabularies and ontologies maintained by ELIXIR-UK in addition to the RDA list. This is also integrated within the DSW

p.46
Verktøy for datahåndteringsplan

We would like to warn against a too high degree of fragmentation of DMP templates, also due to the difficulties of maintainability. We would rather recommend an approach using suggestions on top of one decision tree based template which can build on each other.

https://github.com/EBISPOT/DUO#README by GA4GH is an existing solution for machine readability of data sharing/access/reuse restrictions from consents

p.47

Det kan eventuelt vurderes større grad av samarbeid mellom institusjonene om (anskaffelse av) elektronisk labjournal.

See above; We would like to stress that the solutions provided within the commercial ELN landscape are mostly not compatible with the existing metadata standards and the focus should be rather open source solutions with support for ontologies.
The chosen solution has to support provenance tracking (in line with R1.2)