

FAIR data management in molecular life sciences

AUTHORS (ALPHABETIC ORDER)

FATEMEH Z. GHAVIDEL, COMPUTATIONAL BIOLOGY UNIT, DEPT. INFORMATICS, UNIVERSITY OF BERGEN
Email: Fatemeh.Ghavidel@uib.no

INGE JONASSEN, DEPT. INFORMATICS, UNIVERSITY OF BERGEN
Email: inge.jonassen@ii.uib.no

RUNE KLEPPE, COMPUTATIONAL BIOLOGY UNIT, DEPT. INFORMATICS, UNIVERSITY OF BERGEN
Email: rune.kleppe@uib.no

CHRISTINE STANSBERG, COMPUTATIONAL BIOLOGY UNIT, DEPT. INFORMATICS, UNIVERSITY OF BERGEN.
Email: Christine.Stansberg@uib.no

Researchers are now asked to produce a Data Management Plan as part of a funding application to the Research Council of Norway or to H2020. But, what do we mean by data management, and how can this be implemented in your research project in a practical way?

Data management and the FAIR guiding principles

This new development comes as a result of the increasing amounts of data generated within the life sciences. It is becoming a struggle to process and manage data at the same pace as it is produced. At the same time the potential and benefits of publicly funded research data is not fully exploited, simply because the results may not be possible for others to find or use. The increased focus on open research data for publicly funded research leads to an increased awareness on the importance of data management. Open data implies availability for use by others, for a long period of time, in a way that enables others to understand the data and re-use them. This soon becomes an issue of how data is managed in the research project; such as how the data is generated, the analysis is performed, which data formats are being used and how the data is archived and deposited. When applying for fund-



ing, you are now therefore asked to describe a plan for how you will handle these issues in your research project and make the data and results available, or for sensitive data, how you plan to keep them safe. Such a description is referred to as a data management plan (DMP). Both the National e-Infrastructure for Research Data (NIRD) and the Norwegian Open Research Data Infrastructure (NORDi) provide

general support tools to make a DMP for your project [1,2].

An international consortium has established a set of guiding principles – the FAIR principles – that set out to make research data and related metadata discoverable, accessible and understandable today, tomorrow, and well into the future (see Box 1) [3]. FAIR data management refers to principles that

support open data by being Findable, Accessible, Interoperable and Reusable. FAIR management of data requires thorough documentation and proper annotation and addition of metadata (description of the data and how it was generated) throughout all steps in a research project (see Box 2). For researchers it is also important that FAIR management of data is convenient to implement, typically through available tools and/or infrastructures tailored for this.

Where to deposit your data?

ELIXIR, the pan European bioinformatics infrastructure, is a key e-infrastructure for biological and medical research data, and has a central position on the ESFRI roadmap [4]. On the European scale, ELIXIR has identified a set of data resources of fundamental importance to the wider life-science community and the long-term preservation of biological data [5]. This concept is also pursued on a global level, and a global coalition for life science data sustainability is currently working towards securing sustainable core data resources that are important globally [6]. For easy

Sciences (NeLS: <https://nels.bio-info.no/>), for easy analysis, storage and sharing of data. NeLS allows Norwegian scientists and their collaborators (also abroad) to store and share data generated in projects as well as to analyse them using provided bioinformatics tools and workflows. The storage facility uses the National e-Infrastructure for Research Data (NIRD), run by UNINETT/Sigma2, which allows archiving of data. The data analysis workflows are offered as part of the

management to facilitate FAIR data and model management for research projects. SEEK is particularly adapted towards systems biology, but can be used in a general way. SEEK organises the content according to the ISA format; Investigation, Study, Assay; to structure the data and metadata. It is flexible in terms of the type of data stored. Data can be pointed to the local data storage while the metadata or smaller derived data sets can be kept in SEEK. In addition SEEK also

Box 2. How to make your research FAIR

- Use standard formats
- Supply complete metadata (e.g. <https://FAIRsharing.org>)
- Embrace Ontologies
- Use persistent and unambiguous identifiers
- Put your data in a long term stable repository
- Cite, share freely and encourage others

bioinformatics services from ELIXIR Norway.

FAIRDOM and the SEEK platform
The SEEK platform (<http://seek4-science.org/>) within the FAIRDOM

provides several tools to implement standards in DM, e.g., COPASI and JWS, softwares for simulation and analysis of biochemical reactions and networks and Rightfield, a tool for annotating spread sheets to create semantically aware Excel spreadsheet templates.

Box 1 The FAIR Guiding Principles [3]

- Findable: data is human/machine readable and attached to persistent identifiers
- Accessible: data can be found and retrieved by humans/machines using standard formats
- Interoperable: data can be exchanged and used between systems
- Reusable: data can be understood and used by others

reference, the journal Scientific Data has further published a list of recommended data repositories meeting the requirements for FAIR data access, preservation and stability, on their webpage [7] (see Table 1 for examples).

ELIXIR Norway and NeLS

Several national research infrastructures are important for storing, managing and computing on data. The Norwegian node of ELIXIR has developed the Norwegian e-Infrastructure for Life

Hub (www.fair-dom.org) provides a rich support of metadata and data

management, DLN has collaborated with the bioinformatics in-

Table . A selection of Data repositories for long term archiving of biological data

Data type	Short name
Nucleotide sequences	ENA
High-throughput functional genomics experiments	ArrayExpress
Human genetic and phenotypic data	EGA
Protein sequences	UniProt
Mass spectrometry-based proteomics data	PRIDE
Protein-protein interactions	STRING-db
Bioactive drug-like small molecules	ChEMBL
Enzyme and enzyme-ligand information	BRENDA
Metagenomics data	SILVA

frastructure ELIXIR Norway and the FAIRDOM Hub at the University of Manchester. Together they have integrated NeLS and SEEK to enable SEEK referring to large data files residing within NeLS, typically using SEEK to store metadata and systems biology models (if relevant) and NeLS to store the associated omics data. This makes it easier for Norwegian researchers and their international collaborators within the life science field to do FAIR data management, by allowing access to Norwegian infrastructure for storage, computing and archiving of

research data inside NeLS as well as a rich toolbox for FAIR project data management provided by SEEK.

References

- [1] <https://easydmp.sigma2.no>
- [2] <http://www.nsd.uib.no/arkivering/datahanderingsplan.html>
- [3] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship (2016) *Sci. Data* 3:160018

[4] <https://www.esfri.eu>

[5] <https://www.elixir-europe.org/platforms/data/core-data-resources>

[6] Anderson, W.P.A global coalition to sustain core data (2017) *Nature* 543(7644): 179

[7] <https://www.nature.com/sdata/policies/repositories>

FEBS Network

FEBS Network er et tiltak fra FEBS som har til hensikt å støtte forskere innen molekylær- og cellebiologi. Fra nettsiden network.febs.org kan man blant annet finne de fire temakanalene: Early Career Scientist, Educator, Viewpoints og Research. Dersom man for eksempel tar en titt på kanalen Educator, finner man invitasjoner til workshops og seminarer i tillegg til nyttige knep og verktøy for bruk i undervisning. Man kan registrere seg som bruker, noe som åpner muligheten for å lage ens egen profilside. Her kan man presentere sine forskningsinteresser og legge inn dokumenter, bilder og annet. Det er muligheter for å starte emnespesifikke rom, hvor forskningsprosjekter, samarbeid mellom forskjellige grupper og annet kan diskuteres. De forskjellige biokjemiske selskapene har sine sider i systemet, og for eksempel blir NBS Kontaktmøte 2019 presentert på en fremtredende måte.

Nettverket er foreløpig ungt og uferdig, men det er mange utviklingsmuligheter i systemet, og NBS vil anbefale alle å ta en titt på hjemmesiden til nettverket, registrere seg og ta nettverket i bruk.

Generalsekretæren

Skal du arrangere et vitenskapelig møte?

Mulighet for gratis annonsering i NBS-nytt. Dersom NBS-nytt har ledig plass så tilbyr vi gratis annonsering av vitenskapelige møter som er av interesse for medlemmene.

Kontakt en av de nye redaktørene Erik Boye (Erik.Boye@rr-research.no) eller Kristian Prydz (kristian.prydz@ibv.uio.no) dersom du ønsker å benytte deg av denne muligheten.



ORD KLØVERI VED ERIK BOYE

Denne spalten skal ta for seg en del norske språkblomster, kanskje med hovedvekt på de som blomstrer i overgangen mellom vitenskapelig og mer folkelig bruk. Overføringen av ord fra forskning og vitenskap til dagligtale resulterer ofte i bruk som påkaller både gråt og latter. Og noen ganger tar denne overgangen fra oss gode, formålstjenlige ord ved å tillegge dem helt nye betydninger. Når den nye betydningen virker fullstendig meningsløs og ulogisk, får jeg problemer og må skrive

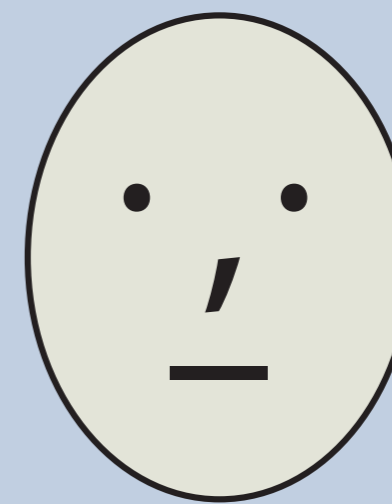


frustrasjonen av meg. I denne spalten vil vi høytidelig og med litt humor minnes ord vi savner. Å bekjempe den nye bruken, er som å slåss med vindmøller, men det hjelper på sjelefreden å skrike litt. Planen er å ha en liten spalte i hvert nummer av NBS-nytt. Men redaktøren av spalten er svært åpen og mottakelig for supplerende innlegg og debatt. Skriv til meg på eboye@rr-research.no.

Komma

Min bestemor lærte meg reglen "punktum-punktum-komma-strek, det var hele Eriks fjes". Det var min første kommaregel, eller kommaregle om du vil. Se figur som jeg har tegnet med Påvverpøint. Siden fulgte mange flere kommaregler, ofte til hjelp og nytte, men mange ganger til forargelse. Jeg har vurdert å kjøpe meg et domene på den store Internett og da fant jeg en instruksjon på hvordan jeg kunne kjøpe, for eksempel, "comma-regel.com". Hvis jeg gjorde det, fristet de med at jeg kunne få epostadressen kanikke@commaregel.com. Det ville vært noe, det! Eksempelet fortalte meg at det er fler enn meg som tenker på kommaregler. Derfor har jeg fått vann på mølla, i alle fall nok til å drive hjernen noen minutter.

Klassikeren på feltet er Per Egil Heggens boktittel om "Heng ham ikke, vent til jeg kommer". En flytting av kommaet mot venstre vil være fatalt for objektet, som da ville blitt eliminert (som det heter i militær sammenheng). Dette er selvfølgelig et oppkonstruert eksempel, men det finnes mange eksempler i dagliglivet der kommaet kan redde om ikke livet, så i alle fall meningen i setningen. Kommaet er en liten tass som kan få stor betydning. Jeg tenker da ikke på kommafeil i tall, som er en helt annen historie. Diskusjonene rundt Statsbudsjettet i



skrivende stund dreier seg ikke om de tusen milliardene, men om småpengene ned mot komma-plaseringen.

Det er ikke lenge siden jeg så en annonse for en leilighet som hadde "et lite sjarmerende kjøkken". Et komma der ville gjort susen; det er nok mer attraktivt med et "lite, sjarmerende kjøkken". I den pågående diskusjonen om hva Kr.F. bør velge av samarbeidspartnere, stod det i en herværende, stor avis at Kr.F. er et lite populært parti. Intensjonen med setningen var nok slik at det ville passet med et komma.

Kommareglene på norsk er ikke så vanskelige og de hjelper oss til å skille forskjellige ledd fra hverandre. På engelsk er det mye verre og krever litt mer Fingerspitzengefühl. Vi

som skriver såpass mye på engelsk, bør kanskje bruke litt tid på å lære oss grunnreglene. Eller – det er kanskje riktig å si at de knapt har strikte regler, men heller en slags konsensus for bruk av komma. Det er for eksempel vanlig at når man flytter fram en setningsdel som normalt hører til lenger bak, bør den skilles ut med et komma. For eksempel: "For this purpose, we examined the cells in a microscope". "Before irradiation, the cells were washed in PBS". En slik bruk av komma er vanlig, men ikke fullstendig gjennomført. På engelsk brukes komma ofte for å lage en liten pause i en lang setning. Trenger du en puste- eller tenkepause, kan du legge inn et komma på et passende sted. Når du leser engelske tekster, ha øynene åpne for hvordan kommaet brukes i praksis. Men jeg er redd du ikke kan velge deg ut en hvilken som helst naturvitenskapelig publikasjon, for denslags forskere er vel ikke akkurat kløppere i feltet. Til slutt vil jeg slå et slag for punktumet. Bruk det. Korte setninger er ofte glimrende. Konsise. Lette å lese. Gir en huggende rytme. Gjerne uten subjekt og verbal. Det har blitt populært nå.