UIT GOES OPEN

Et festlig skrift til Stein Høydalsvik

RAVNETRYKK

UiT Universitetsbiblioteket

RAVNETRYKK

UiT Norges arktiske universitet – 2020

Telefon: +47 77644000

E-post: postmottak@uit.no

Web: https://uit.no/

Septentrio Academic Publishing

https://septentrio.uit.no/

Ravnetrykk, nummer 39 (2020)

ISBN 978-82-91378-47-3

ISSN 2535-8316

DOI: https://doi.org/10.7557/rt.39

UiT goes open – innledningen til et festlig skrift

Per Pippin Aspaas, Aysa Ekanger og Johanne Raade

Sammendrag

Vårsemesteret 2020 runder Stein Høydalsvik 67 år. Samtidig går han over til en tilværelse som pensjonist, etter å ha brukt 23 av sine yrkesaktive år i Universitetets og Universitetsbibliotekets tjeneste. Med Steins hjelp er UiT Norges arktiske universitet blitt en spydspiss innen åpen vitenskap. Mye som i dag er standard i sektoren, ble først pilotert ved UiT under Steins kyndige veiledning. Tjenester innen tidsskriftpublisering, forskningsdatakuratering, master- og doktorgradstilgjengeliggjøring er implementert og raffinert i senere år. En internasjonalt anerkjent konferanse om vitenskapelig publisering er etablert. Fond for åpen publisering, aktiv åpen arkivering av vitenskapelige artikler, en internasjonal, gratis søketjeneste for åpent tilgjengelig forskning om det polare, med mer. Det er få mennesker som evner å sette så markante spor etter seg som Stein. Ikke at han har gjort alt dette selv, men bak svært mye som har foregått ved UiT og ved Universitetsbiblioteket især, har Stein vært en initiator eller inspirator. Det er hans livsverk som UiT-ansatt vi feirer med dette skriftet.

Tempora mutantur …

Verden forandrer seg. Ikke all forandring er av det onde. Det finnes folk som ønsker det beste for oss alle. Slike som står på hver dag for å gjøre verden ørlite grann bedre. Noen slike «endringsagenter» inntar talerstoler, påtar seg verv, skaffer seg plattformer der de kan påvirke fra toppen. Andre arbeider mer i det stille, bak kulissene, med detaljene, med tjenestene, med verktøyene som skal til for at fagre løfter og visjoner skal bli virkelighet. Slike tilretteleggere tildeles sjelden titler eller medaljer. De inntar ikke podier, fyller ikke leserinnleggspaltene. Men de er minst like viktige.

Dette feststemte skriftet er tilegnet en slik endringsagent, Stein Høydalsvik. I 1997 kom han til Universitetsbiblioteket i Tromsø, med hovedfag i astrofysikk og jobberfaring innen IT-bransjen på CV'en. Han valgte jobben som IT-leder, avdelingsleder, faggruppeleder og seniorrådgiver på et offentlig bibliotek, et samfunnsoppdrag han hadde sans for. Kjernen i et universitetsbiblioteks oppdrag er å stille kunnskap og ressurser til disposisjon for alt og alle, gratis.

https://doi.org/10.7557/15.5513

Noe av det første Stein tok fatt på, var å igangsette prosjektet «Det digitale biblioteket». Midt på 1990-tallet var dette fremdeles en ny tanke. Teknologien var der, internettet var begynt å forgrene seg inn i de tusen hjem. Men universitetsbibliotek var fremdeles primært det som uttrykkes i det greske navnet *biblio-theke*, oppbevaringsplasser for bøker. At bibliotekene skulle få en sentral rolle i å spre kunnskap via andre formater enn det trykte, var fremdeles en luftig visjon. Ikke en endring av det onde nødvendigvis, men en endring som krevde strategisk tenkning i kombinasjon med mot. Mot til handling, utprøving, testing, pilotering. Små prosjekter for å bygge kunnskap, skaffe erfaring. Det har hele veien vært Steins metode. Begynne i det små. Se om ting fungerer. Hvis resultatet blir dårlig, forkaste og gå videre. Hvis det blir lovende, ta vare på det, bygge det ut, lansere, implementere, raffinere.

Det digitale biblioteket ble en suksess. I 1999 vedtok Universitetsbiblioteket i Tromsø, som det første i Norge, at tidsskrift i digitalt format skulle prioriteres framfor trykte tidsskrifthefter. Det digitale biblioteket skulle vokse, det trykte formatet var ikke framtidsrettet nok. Beslutningen kan synes selvsagt i dag. Men den gang var det et modig vedtak. Og det har aldri vært i nærheten av å bli forkastet. Snøballen har bare rullet én vei.

E-bøker framfor trykte bøker var en ny milepæl, som UiTs universitetsbibliotek landet på i 2011. Innkjøpspolitikken ble lagt om på følgende måte: Dersom det var mulig for institusjonen å kjøpe en e-bok i et akseptabelt format (helst PDF, uten krav om innlogging), skulle det elektroniske formatet foretrekkes. Oppbevaringsplassen for bøker skulle være «først og fremst digital», slik det het i et av strategidokumentene til Universitetsbiblioteket. Det gjaldt å legge til rette for sømløs tilgang til kunnskap, 24/7, for alle vitebegjærlige.

E-bok- og e-tidsskrift-strategiene fungerte i hovedsak fint, men blottla samtidig et par dilemmaer. Elektronisk litteratur som kjøpes inn av et universitetsbibliotek, kan for det første ikke lånes ut til enkeltmennesker som ikke er studenter eller ansatt. Av rettighetsårsaker må det bygges brannmurer mot samfunnet utenfor. For det andre kan e-bøker ikke lånes ut til andre bibliotek. Fjernlånsordningen, som har tjent brukerne av landets bibliotek i uminnelige tider, blir det satt en stopper for. En papirbok som et bibliotek ikke har anskaffet selv, kan lånes inn fra en annen institusjon. Med en e-bok er dette som regel ulovlig. Kunnskapen blir dermed institusjonsintern. Samfunnsoppdraget, å bidra med kunnskap til alle, stikkes det kjepper i hjulene for.

Politikken med elektroniske tidsskrift og elektroniske bøker har også et kostnadsaspekt ved seg. I lys av teknologien som nå finnes, er det absurd at offentlig ansatte forskere skal gå til kommersielle aktører for å få delt sin kunnskap. Kommersielle aktører som låser kunnskapen inn og selger den tilbake til universitetene for stadig høyere summer, med stadig strengere restriksjoner mot deling innad og

utad. Kunnskapen bør isteden være åpent tilgjengelig, et offentlig gode for alle, hevder *Open Access*-bevegelsen.

Steins sjef i hele perioden fram til 2014 var bibliotekdirektør og historieprofessor Helge Salvesen. Denne historikeren var ikke fremmed for å endre historiens gang. På et seminar Universitets- og høgskolerådet arrangerte i Oslo i november 2003, drøftet Helge det nye fenomenet Open Access, åpen tilgang. I 2003 var begrepet nettopp blitt definert og lansert internasjonalt. Helge peker ut en vei videre for norsk forskning, i tråd med den seneste utviklingen i internasjonal tenkning rundt tilgjengeliggjøring og spredning av kunnskap. Foredraget ble den gang lagt ut åpent på UHRs nettsted, som seg hør og bør. Men når vi i forbindelse med dette Ravnetrykket skulle hente teksten fram igjen, var foredragsmanuset forduftet. Først etter en del plunder lot det seg finne, og da ikke ved hjelp av Google eller andre standard søkemetoder. Det gjenpubliseres her, komplett med DOI (digital object identifier), en varig lenke som skal gjøre publikasjonen motstandsdyktig mot framtidige endringer av UiTs nettsidestruktur.

Situasjonen med Helges forsvunne foredrag illustrerer et viktig poeng: Hovedsaken med åpen tilgang er at forskningsresultater skal legges på nett gratis for alle, og med en friest mulig lisens som åpner opp for viderebruk til for eksempel undervisningsmateriell eller i antologier. Imidlertid er ikke gratis tilgang og åpen lisens nok, det må også bygges infrastrukturer som kan ta vare på det digitale materialet og knytte det sammen med annet innhold på nettet. Lenker må være klikkbare, også om et år eller femten. UHRs nettsider fra 2003 var ikke rustet for dette, og det er dessverre tilfellet med det aller meste som legges ut på nettet også i dag, gratis eller ikke.

Oppbygging av funksjonell infrastruktur for åpen deling, oppbevaring og gjenfinning av kunnskap – og det å bidra til forståelse for viktigheten av dette – er Steins livsverk som UiT-ansatt. Det er det livsverket vi nå feirer.

… nos et mutamur in illis

Både Universitetet og Universitetsbiblioteket i Tromsø har endret seg siden 1990-tallet. Etter en rekke fusjoner er navnet nå UiT Norges arktiske universitet, med filialer – og egne biblioteklokaler – i Harstad, Narvik, Hammerfest, Alta og sju steder på Tromsøya. Den geografiske spredningen alene er et argument for sterk digital kompetanse og infrastruktur for samhandling internt. Men UiT er ikke bare en arbeidsplass der en til hverdags snakker sammen via digitale verktøy. Det er også en institusjon som satser strategisk på *Open Science*, eller åpen vitenskap.

Tittelen på denne antologien, som et knippe av Steins kolleger har bidratt til, er *UiT goes open*. Etter at Helge Salvesen gikk av med pensjon i 2014, overtok historiker Johanne Raade bibliotekdirektørstolen. I sin artikkel diskuterer hun det som har skjedd siden Helges foredrag fra 2003. Er forskningen blitt så åpen som idealistene, også

her ved UiT, håpet på ved inngangen til det nye årtusenet? Fremdeles gjenstår en hel del. Men en skal ikke underslå betydningen av de mange små skrittene. Skal endring oppnås, må berørte parter – først og fremst forskerne og underviserne selv – overbevises. Og skal de overbevises, må de som skal snakke til dem og med dem, ha kunnskap om detaljene og dilemmaene. Først da kan vi si som dikteren: «Tidene endrer seg, og vi endrer oss også med dem».

To eksempler på slike aktiviteter myntet på kunnskapsoppbygging og holdningsendring er den årlige *Munin-konferansen* (The Munin Conference on Scholarly Publishing), som startet opp helt tilbake i 2006, og podkasten *Open Science Talk*, der den første episoden kom i 2018. Leif Longva dokumenterer i dette Ravnetrykket konferansens historie, diskuterer tenkningen som lå til grunn for oppstarten, og legger ved en komplett liste over hovedforedragsholdere fra 2006 til 2019. Konferansen begynte som en norskspråklig møteplass, men gikk etter få år over til å bli en rent engelskspråklig konferanse, med gjenklang langt ut over landegrensene. Gjennomslag internasjonalt er også solid dokumentert i Erik Lieunghs bidrag om Open Science Talk. Det aller første intervjuobjektet i podkastserien var ingen ringere enn Stein Høydalsvik, som fortalte om fenomenet åpen vitenskap i hele sin bredde. De 33 episodene har vært nedlastet ofte, og mer enn halvparten av lytterne befinner seg utenfor landets grenser. Det er grunn til å tro at det er en hel del konferansedeltakere og podkastlyttere som gjennom sitt møte med kompetansemiljøet rundt Stein Høydalsvik er blitt (enda ivrigere) ambassadører for den åpne vitenskapen.

Kunnskap og holdninger er essensielt, men like viktig er det å legge til rette for nye arbeidsformer for forskere som ønsker å praktisere åpenhet. Johanne Raade nevner i sin tekst forskningsdata som et av flere satsingsfelt ved UiT de siste årene. Den spede begynnelsen kom med *TROLLing*, The Tromsø Repository of Language and Linguistics, i 2014. Bakgrunnen var en erkjennelse av at forskningsdataene som lå til grunn for vitenskapelige artikler og bøker var minst like viktige som publikasjonene selv. 87 datasett har funnet veien fra forskernes computere til TROLLing siden oppstarten. Mange års arbeid med språkdata kan her lastes ned, granskes og bygges videre på. Men der det finnes flust med retningslinjer for hvordan en skal sitere artikler og bøker, er oppskriftene på hvordan forskningsdata skal siteres og krediteres en mangelvare. Helene N. Andreassen forteller om hvordan hun med basis i sin erfaring som kurator for TROLLing-arkivet og underviser i forskningsdata ved UiT har vært med på å formulere et sett med retningslinjer som nå brukes av språkforskere verden over, The Tromsø Recommendations for Citation of Research Data in Linguistics.

Enorme mengder data som kunne vært lagret og delt åpent, går tapt hvert eneste år i forbindelse med at forskere går av med pensjon eller flytter. I 2016 ble *UiT Open Research Data* lansert, som en tjeneste for alle forskere ved institusjonen, uavhengig av fagfelt. Lars Figenschou forteller i sin artikkel om *Seniordataprosjektet*, der målet har

vært å samle inn, digitalisere og publisere data fra de av UiTs forskere som nærmer seg slutten av karrieren. I prosjektperioden, 2017–2019, er det samlet inn og lagt ut flere hundre datasett fra denne målgruppen, innen alle vitenskapsgrener fra breddeuniversitetet UiT.

I 2017 ble egne retningslinjer for forskningsdata ved UiT lansert. I kjølvannet av retningslinjene har forskningsdata-arbeidet ved UiT tatt stadig nye steg. *DataverseNO* er lansert, et samarbeid mellom til sammen ni involverte institusjoner. En av dem som har jobbet tett sammen med Stein med oppbyggingen av dette nasjonale arkivet for åpne forskningsdata, er Philipp Conzett. I sin artikkel skriver han om de mange små og store grepene som skal til for å sørge for at forskningsdata skal følge FAIR-prinsippene (Findable, Accessible, Interoperable, Reusable). Forskningsdata er komplekst. Det er ikke bare å trykke på en knapp der det står «save and publish». Før en kommer så langt, er det et mylder av tekniske og fagspesifikke utfordringer må tas stilling til for å gjøre åpne forskningsdata så FAIR som mulig.

At noe er åpent tilgjengelig og trygt lagret, betyr ikke nødvendigvis at det er lett å finne. Svært mye åpent forskningsmateriale, ikke minst forskningsdata, ligger på servere som ikke «snakker med» resten av internettet. Verdifulle data fra en lang rekke forskningsinstitusjoner må en klikke seg langt inn i nettsidestrukturen til den aktuelle institusjonen for å finne. For å bruke FAIR-begrepet: slike data er kanskje Accessible, men neppe Findable, og da er en like langt. Prosjektet *Open Polar*, ledet av Tamer Abu-Alam, handler om å bygge en søketjeneste som finner fram til så mye av den åpent tilgjengelige kunnskapsproduksjonen om det polare som overhodet mulig. Polarforskere fra både UiT og Norsk polarinstitutt er med som rådgivere underveis. Prosjektet bygger for øvrig på et annet prosjekt som Stein var sentral i å bygge opp, søketjenesten High North Research Documents, som ble lansert i 2013. Med Open Polar raffineres søketjenesten og utvides betydelig både teknisk og innholdsmessig.

Forskning, forskning, forskning. Det måles og veies og telles. Jan Erik Frantsvåg og Per Pippin Aspaas er i sin artikkel opptatt av å telle artikkelpubliseringen ved institusjonen. Hvor mange artikler publiseres i rene Open Access-tidsskrift? Hvor mange fagfellevurderte artikkelmanus arkiveres i det åpne vitenarkivet Munin? Og hvor mye koster det å forvalte et fond for åpen publisering, til dekking av APC (article processing charges)? Bakteppet for deres artikkel er UiTs policy-vedtak i 2010, der publisering i åpne tidsskrift og arkivering av artikler i Munin ble framhevet som ønskelig. I 2019 ble denne policyen gjort mer allmenngyldig, parallelt med at forskningsadministratorer fulgte opp forskerne tettere enn før. Statistikken for åpen publisering viser en rask marsj oppover fra 20 % åpne artikler i 2011 til godt over 80 % i 2019.

Et annet tiltak kom på nasjonalt plan, de såkalte «publiser og les-avtalene», eller *Unit-avtalene*, som er tegnet med flere av verdens største kommersielle utgivere av vitenskapelig litteratur. Takket være

disse nasjonale avtalene, som begynte å virke fra 2019, blir artikler som er skrevet av forskere tilknyttet norske institusjoner, åpent tilgjengelige nærmest automatisk i en lang rekke anerkjente tidsskrift. UiTs erfaringer med avtalene så langt belyses i en egen artikkel. Det er talende for veksten i åpen publisering at artikkelens forfatter, Lene Ottesen, til daglig arbeider med innkjøp av litteratur. Der åpen tilgang tidligere var noe en engere krets i en egen avdeling i biblioteket arbeidet med, har det nå vokst seg til å bli en del av hverdagen for mange flere. Lene viser i sin artikkel til kostnadene den formen for «hybridpublisering» som Unit-avtalene medfører. Hun stiller spørsmålet om ikke publiseringen av vitenskapelig litteratur i større grad kunne tas over av de offentlige institusjonene selv.

I skyggen av de store debattene om kommersielle forlagsmodeller og internasjonale initiativ som Plan S, har forskningsbibliotek verden over bygget egne publiseringstjenester for vitenskapelige tidsskrift. Den åpne programvaren Open Journal Systems ligger til grunn for UiTs *Septentrio Academic Publishing*, som i løpet av få år har vokst til å bli hjem for et tjuetalls åpent tilgjengelige serier. Blant disse er det flere som er primært rapportpreget eller formidlingsrettet, slik som Ravnetrykk. Andre er tradisjonelle fagfellevurderte tidsskrift. I sin artikkel gjør Aysa Ekanger og Solveig Enoksen rede for det arbeidet med å legge til rette for gode rutiner for fagfellevurdering og annen kvalitetssikring av publikasjonene i Septentrio. Publiseringstjenesten er et gratis tilbud for fagmiljøene. Den tilbys med svært små personalressurser, men leverer høy kvalitet med hensyn til langtidsbevaring og gjenfinning i internasjonale søkemotorer. Sammen med bibliotekets to IT-ingeniører, Karl Magnus Nilsen og Obiajulu Odu, planla Aysa i dette nummeret av Ravnetrykk å virkeliggjøre en visjon Stein har hatt lenge, nemlig at Septentrio-tidsskrift ikke bare skulle komme ut i formatet PDF, men også i andre formater som XML, HTML og EPUB. Forsøket med å XML-kode dette Ravnetrykket strandet imidlertid helt på målstreken. I en egen redegjørelse forklarer Obiajulu og Aysa hvorfor de mislyktes. Artikkelen er et eksempel på at ikke bare suksesshistorier fortjener å publiseres. Godt, transparent FoU-arbeid bærer i seg en forpliktelse om å være åpen om mislykkede så vel som vellykkede forsøk. Få ting er så bra som å kunne lære av andres feil.

Transparens er også noe av hovedinnholdet i en debattartikkel skrevet av Bård Smedsrød og Leif Longva. De diskuterer behovet for å bevege seg vekk fra en prestisjefiksert publiserings(u)kultur mot en kultur preget av rask, ærlig og åpenhjertig deling av forskningsresultater. Vi er her tilbake til et av de store dilemmaene de to bibliotekdirektørene tar opp i sine bidrag. Likevel, om så alle forskere begynte å diskutere og publisere åpent heller enn lukket, ville det ikke være nok. Kunnskapen må også formidles. Nye generasjoner av samfunnsborgere må lære seg å finne vei i informasjonsjungelen. Mariann Løkse forteller i sin artikkel om erfaringene med Universitetsbibliotekets nettkurs for ferske studenter, *iKomp*, et klassisk eksempel på en åpen undervisningsressurs (OER, Open

Educational Resources). Dette digitale undervisningstilbudet består av tekster, videoer og test deg selv-oppgaver og gis både i kombinasjon med og løsrevet fra tradisjonell klasseromsundervisning. En videreutvikling for elever i videregående skole ble lansert i 2019, utviklet av UiT i samarbeid med Troms fylkeskommune.

Samarbeid med andre aktører både utenfor og innenfor universitetet ligger forresten til grunn for flere av de tjenestene Stein har vært med på å bygge opp. Det er mye som kunne vært nevnt. Nettressursen *Arkitekturguide for Nord-Norge og Svalbard*, som ble til som et samarbeid mellom Universitetsbiblioteket, kunsthistorikermiljøet ved universitetet og Arkitektforeningen i landsdelen, har i perioder vært en av UiTs mest besøkte nettsider. *Munin*s ferd fra en liten DSpace-installasjon for deling av utvalgte studentoppgaver («Electronic Theses and Dissertations»-prosjektet) til obligatorisk innleveringsportal for samtlige fag ved hele UiT er en annen begynne-i-det-små-historie som har vokst til en tjeneste til glede for de mange.

Med sin infrastruktur for tilgjengeliggjøring av forskningspublikasjoner, forskningsdatasett og undervisningsressurser er UiT virkelig i ferd med å bli åpent.

Gratiarum actio

Til sist gjenstår det bare for oss redaktører å takke – og beklage. Takk til alle som har bidratt, med velskrevne bidrag innenfor knappe tidsfrister. Takk til Linnea Nordström for arbeid med grafisk uttrykk for Ravnetrykk og for design av forsiden til dette nummeret. Og, ikke minst, beklager til alle dem som helt sikkert gjerne skulle skrevet noen ord, men som ikke har blitt invitert med. Vi har rett og slett vært så opptatt av å holde prosjektet hemmelig, at vi kun har «headhuntet» noen ytterst få skribenter. Men selv om prosjektet *UiT goes open* har vært langt fra åpent, er resultatet hermed lagt ut med åpen tilgang for hele verden, som så mye annen kunnskap fra UiT.

Tromsø, 31. mai 2020

Per Pippin Aspaas, faggruppeleder

Aysa Ekanger, Open Access-rådgiver

Johanne Raade, bibliotekdirektør

Addendum

Tempora mutantur, nos et mutamur in illis er et sitat som ofte tilskrives dikteren Ovid. For mer enn to tusen år siden skrev han, i diktverket *Fasti*, noe litt annet (bok VI, vers 771–772):

> *Tempora labuntur, tacitisque senescimus annis*
> *et fugiunt freno non remorante dies.*
>
> Tidene svinner, og alderen kommer med årene stille.
> Dagene flykter, det fins ikke en tøyle for dem.

Dette er så blitt endret gjennom tidene, inntil kreative latindiktere fra humanismens tidsalder ga det sin nåværende, litt mer optimistiske, form.

Ovids ord lever evig. Men de endres som tiden selv. Ikke av seg selv, men med god hjelp fra kreative hjerner. Små og store endringsagenter som vil det beste for oss alle, i beste humanistiske tradisjon. La oss feire endringene, mens vi unner Stein hans velfortjente *otium*.

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Open Access Initiative – why, and are we willing to try?[1]

Helge Salvesen

Abstract (Editorial)

In the very early days of the Open Access movement (the concept had, in fact, just been coined at that time) former library director at UiT Helge Salvesen held the introductory speech at a seminar called "Open Online Access to Research", organised by the Norwegian Council for Higher Education, in Oslo, November 2003. In his speech, Salvesen describes the forms of Open Access now known as Gold OA and Green OA, emphasising the need for institutional repositories and other kinds of infrastructure to facilitate openness and transparency in research.

Introduction

Welcome to this seminar, which we have called *Open Online Access to Research*. I will go straight to the point and ask: Why do we need an Open Archive Initiative? And why does this initiative come from the library sector? And finally, why is this initiative so difficult to implement?

In my introduction to the seminar, I will ask questions, point to dilemmas, and only tentatively hint at answers. I hope that at the end of the day we will have many answers and good strategies for proceeding.

There are three obvious reasons why the open archive initiative is necessary. First, colleges and universities need to cut costs. One way of doing that is to find some way to counter the rising prices of international scientific literature, especially within the natural sciences and medicine, where prices have risen at a phenomenal rate. And providers are preparing new strategies for pushing prices even higher. The open archive initiative is a possible countermeasure. But will it work, is it sufficient, and is the research community willing to try it?

The second reason for this initiative is that universities and colleges are under increasing pressure to legitimize their activities, both to the

[1] Introductory speech at a seminar organised by the Norwegian Council for Higher Education, Oslo, Norway, November 11, 2003.

https://doi.org/10.7557/15.5500

general public and to the authorities that provide their funding. One way of achieving such legitimacy is to document and publicize the institutions' research and development activities to the outside world. One means that individual institutions can use to publish their own production is an open archive. But this is unlikely to be sufficient on its own as a legitimization and PR tool. Institutions of higher education and research have obligations to society and a duty to disseminate their production which goes beyond what is possible with an open online access system for research dissemination. Nonetheless, this initiative is a step in the right direction. But are the universities and colleges both ready for such a change and motivated to spend resources on it? Are there any incentives that might promote a commitment of this kind?

The third reason is that institutions of higher education need to develop and make available electronic teaching and learning materials for regular and flexible teaching on the internet. An open archive system for teaching purposes at each institution, in which various textbooks, compendia, and other course material are developed and placed in a pool that can be re-used and is freely accessible may be one possible solution. But even if much of this is technically possible, problems of financing and copyright issues will limit the potential. Is there a way of widening the scope of what is technically possible? Can the fact that institutions of higher education need good teaching materials help ensure easier access?

First I will say something about the dilemmas of the open archive as a countermove against the commercial players' publishing systems.

Patterns of scientific publishing

Seeing the consequences of accelerating price increases as a problem solely for the library, is at best a rather narrow view; at worst, it is a perspective which can never solve the problem of high prices for international literature. The dilemma is as much a question of systems for bestowing status and rewards within the scientific community as a technical problem of price negotiation.

The situation can be briefly summarized as follows: There is a great imbalance between academic fields in terms of access to scientific literature. Roughly speaking, literature on the natural sciences, mathematics and medicine is published in international journals, which have an annual price increase of 10–15 per cent. The humanities and social sciences are more monograph-oriented, and their journals are generally somewhat cheaper than natural science and medical journals. In recent years, this has caused a shift in the allocation of university literature budgets in favor of the hard sciences, to the detriment of the humanities and social sciences. One dilemma is therefore that we have to expect different motivations for solving the problem in different institutional departments.

The fact that libraries have prioritized electronic editions of journals has also lead to a decrease in inter-library loans. This is because the

providers of digital journals, i.e., the owners of the information, limit the institutions' right to pass on the information to anybody other than those licensed to use it, i.e., faculty and students at the paying institution.

Since 1 July 2001, literature bought as digital documents is subject to VAT, while paper-based documents are exempt. This automatically gave a 24 per cent price increase for literature available in electronic form. For colleges and universities, an important issue with large economic ramifications is raised when their literature purchasing power is reduced to three-quarters of what it was by the stroke of a pen – or even a key stroke. The decision must be regarded as the result of oversight, but the *Storting* (Norwegian parliament), as the legislative and budgetary authority, has not been willing to remove the injustice.

Why is scientific literature, and journals in particular, so expensive? So far, universities and colleges have been content to regard this as a library problem and note the need for better negotiation strategies at the libraries in order to bring down prices. Something can obviously be done at the purchasing end. Libraries can, for instance, become better at coming together in national, Nordic and international consortia in order to achieve economies of scale. "Pay per view" access is another possible strategy. But such measures would still have only a marginal effect. Furthermore, the large providers keep introducing new strategies so that they can continue to milk the market as much as possible. One way, which has become increasingly common in recent years, is that libraries are given access to a publisher's journals at a package price. And in order to woo libraries and bring down the price per journal, publishers might offer their entire journal portfolio as an extra bonus, even if the libraries do not ask for it. Now we see that publishers are working on plans to charge for all the journals in the package, while still only allowing purchase of access to the entire package. Even if only one of the journals in the package is a must, you have to buy access to the whole lot.

There are two circumstances that make scientific journals a peculiar market: Since the content of the various journals is different, it is not possible to replace an expensive journal with a cheaper one, as is possible with, for instance, a detergent. Furthermore, the production side, i.e. researchers, function in such a way that it is not most interesting to publish in a cheap journal, but rather in the most prestigious ones. Everyone concerned about the future of Norwegian research insists that we must increase our efforts to publish in the leading international journals. Some journals are so prestigious that the authors even have to pay the publisher to have an article printed. Some institutions, also in Norway, have gone so far as to offer extra pay to researchers who manage to get publish in the leading international journals. At the same time, being an editor, referee, or member of the editorial board of the same publications gives high status in the academic community. A natural consequence is that

this system increases the market value of the product which is being quality assured in this manner. The publisher's profit can be increased, since the value added comes at little cost to the publisher. The researchers work mostly for the honor.

In spite of the fact that the academic production and quality assurance are very cheap, the product is expensive. The publisher, whether a university or a commercial publishing house, can see that a high-status international scientific journal is a product with considerable commercial value in the knowledge society. The subscription price can be pushed up. The pain barrier is high, because journals that represent the cutting edge of science must be available in the library. Since the producer, i.e., the researcher, has an interest in the product being as good as possible and is willing to do his very best to achieve this, the publisher will be able to exploit this to the full in a market situation: A top-notch product with low production costs, which the producer himself demands as a customer and will pay almost any price for, makes the market ideal for the publisher. In theory, the price can be increased to the point where the costs of buying back the information become so high that there are no longer funds available to produce the knowledge necessary to make the journal attractive.

In such a market, the libraries have only limited ability to push down prices when buying back for the research community the information that the same research community has been willing to do its utmost to add value to, and which it helps define the ranking of – and thus, to a large extent, also the price. Without a change in the triangular relationship between producer, publisher and consumer, the market forces that are part of this price spiral will mean that only the wealthiest institutions will be able to afford to provide their researchers with the most up-to-date information available. With journals available electronically and with copyright law to hand, publishers will only give access to those who pay for it. There will be no free rides. The market can thus be protected against leaks.

Any potential library-internal solutions to these problems will have major negative consequences. If we choose to cancel our subscriptions to the most expensive journals in order to protect a larger number of cheaper ones, this might mean that important academic fields will disappear completely from the library "shelves". Should we choose to base our priorities on borrowing statistics, this would have equally catastrophic results. What would be the consequences for research if only the richest institutions are willing to pay for access to a rarely used but still scientifically very important journal? We might also be forced to choose between prestigious publications and access to articles intended for students. This is therefore a problem area full of dilemmas of an academic, legal, and financial nature.

Far more actors than just the libraries need to make up their minds about this issue. That is why this seminar is being organized, as a

collaborative project between the library and the research committee of the Norwegian Council for Higher Education. An open archive system based on non-profit making electronic publishing might be a way out of the rut. But until researchers are willing to consider publishing via alternative channels and contributing to a change in publishing patterns, this initiative is doomed to failure. Or is there a possible way out of the vicious circle? Maybe this seminar can provide an answer. A case in point is, for instance, BioMedCentral, a publisher which has created a professional infrastructure for open archive publishing and publishes close to 100 bio-medical journals. Furthermore, since 1991, a number of physicists have been filing their research documents online, both before and after peer-review in the arXiv e-print archive.

Internal publication

Why has internal publication seen a renaissance now, hundreds of years after Gutenberg's invention of printing made it the concern of professional publishing companies? Open access has become relevant, first of all, in relation to *digital* publishing by researchers' home institutions. This kind of publishing has at least three functions: First, scientific institutions welcome internal digital publishing because it saves money now that publishers have turned scientific publishing in particular into a license to print money, – and second, in order to make working papers available to colleagues within the field. Third, institutions use internal publishing to legitimize and publicize their activities to the outside world, but also in order to facilitate access, use and re-use of the material in, for instance, teaching.

There is therefore an international trend in universities and colleges to stimulate internal publication on the web, and there is an increasing emphasis on organizing the information well, storing it safely, and making it easily accessible. Furthermore, institutional libraries are offering students and faculty support for digital publishing, and it is often the libraries that organize this academic production, prepare it, store it, and make it accessible. It is crucially important that this kind of publication should not be open to all; the material must be quality assured academically before being published.

What then is a digital institutional repository? Clifford A. Lynch offers the following description:

> "… a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where

> appropriate, as well as organization and access or distribution."[2]

The MIT Durable Digital Depository has the following objectives for its internal publication strategy: "A repository that captures, stores, indexes, preserves, and redistributes the intellectual output of a university's research faculty in digital formats." In addition to electronic versions of master's and doctoral theses the following could form part of a university's or college's internal publishing services: Articles and reprints, technical reports, working papers, conference papers, posters, various types of data sets such as observation data and data collected for research projects at the institution, visual and scientific images of various kinds, audio files, video files, course materials, and reformatted digital library collections.

Concluding remarks

The question is: Has institutional internal publication been identified as a need in Norway? Are we paying sufficient attention to quality assurance of the content? Are we interested in channeling resources to the libraries for such publication? Are there incentives that can promote such a development? Is it sufficient to point out:

- that the open archive system is one of the very few possibilities we see today of combating the price increases in scientific publication
- that studies show that articles to which there is open access are quoted more often than those only available via subscription
- that the standardization of metadata has made possible the emergence of global providers of search engines that enable searching across a range of digital archives
- that it offers researchers secure long-term storage of their scientific documents
- that the documents will be available on the net and will be searchable in open global catalogues
- that an increasing number of publishers allow parallel publication of researchers' research in their home institutions' electronic archives?

If there is strong enough motivation in a united research community for these new forms of publishing, the tools for success are available: If authors collectively demanded the right to parallel publishing in an open archive system, then the publishers of commercial publications would have to give way, since their publications are totally dependent on the production of the researchers. On granting research funding, the Research Council of Norway, the Ministry of Education and

[2] Clifford A. Lynch, 'Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age', *ARL*, no. 226 (February 2003): 1-7.

Research, and the universities could set the condition that researchers publishing the results of their research reserve the right to add a copy to the electronic archive of their institution. Today's conference might provide an answer to the question whether the motivation for such a change is sufficiently strong.

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Mot åpen vitenskap ved UiT

Johanne Raade

Sammendrag

I denne teksten beskrives kort utviklingen med åpen publisering etter 2003, da Helge Salvesen holdt sitt foredrag om det nye fenomenet Open Access. Salvesen peker i sitt foredrag på noen sentrale dilemmaer i arbeidet med åpen tilgang, slik han så det i 2003. Noen av disse dilemmaene er fortsatt relevante i dagens publiseringsregime. Den gang var det særlig tilgjengeliggjøring av forskningsartikler i åpne arkiv og universitetenes holdning til dette som var bakteppet. I dag vil vi inkludere både åpen publisering og åpne arkiv i det mer omfattende feltet åpen forskning (Open Science), som også inkluderer bl.a. data og undervisning.

Tilbakeblikk

I 2003, da bibliotekdirektør Helge Salvesen holdt sitt innlegg for Universitets- og høgskolerådets to undergrupper UHR-bibliotek og UHR-forskning, ble innlegget trolig møtt med noe skepsis i universitetsmiljøene. Året 2003 var nemlig også året da de fire gamle universitetene sammen begynte å utvikle databasen FRIDA (Forskningsresultater, informasjon og dokumentasjon av vitenskapelige aktiviteter) som et redskap for å registrere publikasjoner som skulle inngå i finansieringssystemet for institusjonene. FRIDA ble innført i 2006 og lever videre som et obligatorisk registreringssystem for publikasjonene til alle forskere i Norge, nå under navnet Cristin (Current Research Information System in Norway). Fokuset på publisering bidro på den ene siden positivt ved at man fikk en samlet oversikt over norsk forskning, en dokumentasjon som tjente som legitimering overfor både myndigheter og befolkningen for øvrig. På den andre siden bidro publiseringsmålingene til et økt fokus på de tradisjonelle publiseringskanalene (tidsskriftene) og på prestisjetidsskriftene i særdeleshet, gjennom et økt press for å publisere i «gode» kanaler. Systemet er kjent som «tellekant-systemet».

Institusjonelle arkiv og debatt om åpen publisering ble av flere oppfattet å være i strid med det nye systemet, og av enkelte også i strid med god kvalitet.

https://doi.org/10.7557/15.5516

Ved universitetene fantes på denne tiden noen engasjerte og entusiastiske forskere og andre ansatte som forfektet at åpen publisering var den rette veien, mange av disse var ansatt i bibliotekene. Bibliotekene var i en konstant pengeknipe på grunn av stadig økende priser på abonnementer og databaser, og var også blant dem som kjente forlagene og de store publiseringshusene best.

Fagbibliotekene hadde begrensede muligheter for å løse utfordringene med de raskt økende prisene, noe som er godt beskrevet i Salvesens artikkel. Åpen publisering ble derfor ansett som en løsning på utfordringene. Bevegelsen startet i de store forskningsbibliotekene i USA og ble plukket opp ved universitetsbibliotek verden over.

Utviklingen ved UiT

Åpen publisering ved UiT har utviklet seg over tid og anslaget i 2003 medførte opprettelse av en plattform for Open Access-tidsskrift fra egen institusjon. Det første tidsskriftet på plattformen ble *Nordlyd*, som ble lansert i heldigitalt format i desember 2003.

Allerede i 2005 startet man en planlegging av et institusjonelt vitenarkiv for UiT og *Munin* ble offisielt lansert 21. september 2006. I denne perioden ble også det såkalte tellekantsystemet lansert og debattene omkring publisering var ikke bare knyttet til åpen publisering, men til publisering generelt. UB tok i sakens anledning initiativ til å etablere en arena for diskusjoner rundt vitenskapelig publisering generelt og den første Munin-konferansen så dagens lys 23. november samme år.

Fra 2008 ble det obligatorisk å levere masteravhandlinger via Munin, avhandlingene ble dermed også arkivert digitalt. Åpen tilgjengeliggjøring av avhandlingene er imidlertid en frivillig sak for studentene og dette har medført at ikke alle avhandlinger er tilgjengelig. De avhandlingene som derimot er tilgjengelige, leses oftere, og blir sitert oftere, enn de som finnes i papirutgave eller de som dessverre ikke er tilgjengelig overhode. I 2013 utvidet Munin sitt nedslagsfelt og det ble obligatorisk å levere også doktoravhandlinger via denne åpne kanalen. Også her ble publisering gjort valgfritt. Mange doktoravhandlinger består av artikler som er utgitt på tradisjonelle forlag, og er derfor underlagt en embargoperiode før de kan tilgjengeliggjøres i institusjonelle arkiv.

Gjennom suksessen med UiTs institusjonelle arkiv ble det etter hvert lansert en ide om å også etablere en tjeneste for åpne tidsskrift og serier, og i 2010 ble *Septentrio Academic Publishing* lansert. Per i dag er det 19 ulike publikasjoner som bruker plattformen, og av disse er 9 registrert som tellende tidsskrift i den nasjonale vitenskapsindeksen, deriblant det nevnte *Nordlyd*.

Én ting er å etablere systemer og verktøy for deling av publikasjoner og tidsskrift. Noe annet er det å forankre disse ideene hos Universitetets ledelse og blant forskere og undervisere. De første forsiktige

skrittene ble tatt av ildsjeler på UB sammen med entusiaster blant forskere på fakultetene og i den sentrale forskningsadministrasjonen, og endelig i 2010 vedtok Universitetsstyret ved UiT prinsipper for åpen publisering som ble gjeldende politikk for hele universitetet.

Et av utgangspunktene for arbeidet med åpen publisering var galopperende kostnader for tradisjonelle tidsskrift, men heller ikke åpen publisering er gratis. Nye publiseringsmåter gav også andre finansieringsmåter for forlagene, og forfatterbetaling ble introdusert. I 2011 etablerte UiT et eget fond for forfatterbetaling. Fondet ble administrert av UB.

Med den økende mengden åpent tilgjengelige forskningsartikler verden over var det i 2012 mulig å etablere en tjeneste som tematisk søkte opp artikler med relevans for nordområdene, og High North Research Documents ble lansert 25. januar 2012 på konferansen Arctic Frontiers. Denne søketjenesten utvikles nå videre under prosjektnavnet Open Polar, som skal inkludere både forskningsartikler, master- og doktoravhandlinger, vitenskapelige rapporter, og ikke minst forskningsdata knyttet til Nordområdene og Antarktis.

Gjennom disse stegene mot åpen publisering hadde UB-ansatte over tid fått kompetanse som ble etterspurt fra fagmiljøene, både som rådgivere innen publisering og som innovative, godt informerte samarbeidspartnere, om det som skjedde på den internasjonale arenaen. På forespørsel fra fagmiljøet i lingvistikk ble det i 2014 lansert et tematisk arkiv for lagring og tilgjengeliggjøring av forskningsdata kalt *TROLLing* (The Tromsø Repository of Language and Linguistics). I arbeidet med utvikling av dette tematiske forskningsarkivet etablerte UB kontakt med internasjonale miljøer som over tid hadde arbeidet med dette feltet. Valget av teknisk løsning falt på den Harvard-utviklede programvaren *Dataverse*, og i 2016 utvidet UB sin portefølje med en egen tjeneste for lagring av åpne data, *UiT Open Research Data*, basert på den samme tekniske infrastrukturen.

I 2017 vedtok Universitetsstyret prinsipper for håndtering av forskningdata ved UiT. Infrastrukturen UiT Open Research Data ble utvidet videre da man samme år lanserte DataverseNO som en nasjonal tjeneste for lagring og arkivering av forskningsdata. Per i dag er det 8 norske institusjoner som har tatt i bruk DataverseNO, i et samarbeid ledet fra UiT.

DORA-erklæringen (The San Francisco Declaration on Research Assessment) ble også underskrevet av UiT i 2017. Denne stadfester i korte trekk at det er kvaliteten på det enkelte forskningsbidraget (publikasjonen/datasettet) som skal vurderes og ikke kanalen/tidsskriftet dette er publisert i. DORA er de seinere år signert av mange institusjoner over hele verden. I Norge har flere institusjoner samt Forskningsrådet underskrevet erklæringen og på denne måten forpliktet seg til å vurdere kvaliteten på forskningen etter de råd og prinsipper som fremkommer av erklæringen.

Kunnskapsdepartementet har også fulgt opp arbeidet med åpen forskning og stiller krav om at all forskning finansiert av offentlige midler skal være åpent tilgjengelig innen 2024.

Ønske om fortgang i åpningen av forskingen førte også til at flere forskningsfinansierende institusjoner kom opp med Plan S. I korte trekk ønsker man kun å finansiere forskning som også gjøres tilgjengelig for alle. Planen er sterkt debattert og flere forskningsmiljøer hevder at dette vil bidra til å stenge norske forskere ute fra internasjonalt forskningssamarbeid og medføre vanskeligheter med å publisere i de ledende fagtidsskriftene. Debatten har bidratt til at problemstillinger rundt åpen publisering nå er diskutert i fagmiljøer og på institusjonsnivå i hele sektoren, og ikke lenger blir ansett som et bibliotekanliggende alene.

Nasjonale og internasjonale dilemmaer

Salvesen trakk opp tre viktige argumenter for et institusjonelt arkiv og åpen publisering i 2003. Har de siste 17 årene bidratt til at vi er kommet lenger?

De tre argumentene var: i) ønske om å kutte kostnader, ii) øke legitimiteten til forskning hos myndigheter og øvrig befolkning og iii) muligheten for tilgjengeliggjøring av materiale til digitale og fleksible studier. Flere dilemmaer ble også trukket frem og er i høyeste grad gjeldende også i dag.

Det vitenskapelige publiserings-markedet er spesielt. Markedet domineres i dag av fem store forlagshus som dekker de fleste fagdisipliner. Det finnes i liten grad alternativprodukter til de anerkjente prestisjetidsskriftene. Tidsskriftet *Nature* kan for en forsker ikke erstattes av andre tidsskrift, verken for lesing eller publisering. Etterspørselen etter disse tidsskriftene er konstant og i liten grad påvirket av prisen. Dette har de store forlagshusene utnyttet ved å kjøpe opp mange tidsskrift og selge tilgang til innholdet i disse i form av stadig dyrere tidsskriftpakker. I løpet av de siste tiårene har eierskap til tidsskriftene beveget seg stadig lengre vekk fra de vitenskapelige institusjonene. Mindre, uavhengige forlag og vitenskapelige foreningers skriftserier er overtatt av kommersielle forlagshus. Men hovedproduktet, altså forskningen, blir fremdeles produsert av forskere på norske institusjoner med i all hovedsak offentlig finansiering. Forskningen er kvalitetssikret av kollegaer på ulike institusjoner med samme offentlige finansiering. Rettighetene til artiklene overgis som regel til forlagshusene ved publisering, og disse kan sette prisen og tilgangsrettigheter til produktet. Innkjøpet gjøres av universitetsbibliotekene, som har begrensede forhandlingsmuligheter.

Hvor står UiT i dag

I UiT's strategi, *Drivkraft i nord*, står det at: «UiT skal være nasjonalt ledende på Open Science og våre forskningsdata og publikasjoner skal være åpent tilgjengelig der dette er mulig».

UiT har altså siden den gryende starten i 2003 tatt steg og ønsker nå å være en tydelig aktør nasjonalt på feltet. Som vist over har utviklingen gått stegvis, men tross dette har UiT i mange sammenhenger vært en foregangsinstitusjon i Norge.

I nasjonal sammenheng har man de siste årene inngått nye avtaler med fire store forlag, såkalte «publiser og les-avtaler». Disse har på kort tid bidratt til at det ikke betales to ganger for en artikkel, ved at forskere som ønsker å gjøre sine artikler åpent tilgjengelig har betalt en avgift til forlaget for å tilgjengeliggjøre den, mens biblioteket samtidig har betalt for abonnementet. I de nye avtalene er det inngått avtale om lesetilgang i kombinasjon med «gratis» publisering for et gitt antall artikler fra Norge. Avtalene er en suksess ved at norsk forskning nå er tilgjengelig for dem som er interessert i å lese den, men man har ikke kommet noen vei med kostnadsbildet. De totale kostnadene for disse avtalene er meget store og det er å håpe at disse «overgangsavtalene» ved neste korsvei kan forhandles ned i pris. Enkeltforskere og forskningsmiljøer gjøres i liten grad prisbevisste gjennom disse avtalene, hvor «åpen publisering» oppleves gratis. Påtrykket fra myndigheter og forskningsfinansiører gjennom særlig Plan S, har bidratt til økt oppmerksomhet og en bred debatt om åpen publisering. Såkalte «røvertidsskrift», som har svak eller ingen kvalitetssikring, bidrar til at enkelte forskere er kritiske til åpen publisering og mener at ting skjer for raskt.

Tilgang til kvalitetssikret forskning er essensielt for tilliten til forskningen og i denne debatten har også tilgangen til data blitt et hett tema de siste årene. Åpen tilgang til publikasjoner og data må gjennom god kvalitetssikring bidra til transparens og etterprøvbarhet og dermed sikres legitimitet. Forskningsbasert kunnskap er et demokratisk anliggende og bør være tilgjengelig for alle som har bruk for den. Til begrepet åpen forskning hører også åpen utdanning, åpen programvare mm. UiT har en helhetlig ambisjon for «åpen» og gjennom sitt arbeid med informasjonskompetanse bl.a. utarbeidet kurset *iKomp* på en åpen plattform (edX). iKomp er gratis tilgjengelig for alle og er utformet i to versjoner: en for universitetets studenter og en for videregående elever. På en digital og åpen plattform tilbys alle som er interessert i det et kurs i kritisk tenkning, vurdering av informasjon til det beste for demokratiet.

Både juridiske, akademiske og økonomiske dilemmaer eksisterer i høyeste grad, i 2020 som i 2003. Rettigheter, kostnader, transparens og tillit til forskning er vesentlig.

Kanskje vi står ved et veiskille etter mange års arbeid. Plan S og debatten rundt forskning, publisering, kvalitet og kostnader er positive signaler som har løftet diskusjonene til nye arenaer.

Nå er tiden inne for å ta diskusjonen om hvor eierskapet til kunnskapen skal ligge. Uten at eierskapet til publiseringen og publiseringskanalene hentes «hjem» til der forskningen produseres, vil vi fortsatt stå i mange vanskeligheter både juridisk og økonomisk.

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Utviklingen i Open Access ved UiT 2011–2019

Jan Erik Frantsvåg og Per Pippin Aspaas

Sammendrag

Dette er ment som en historisk oversikt over utviklingen i viktige tall på Open Access-området (OA, dvs. åpen tilgang til vitenskapelige publikasjoner på nettet) fra og med 2011 til og med 2019. Universitetet i Tromsø, nå UiT Norges arktiske universitet, vedtok sin første OA-policy i universitetsstyremøtet i oktober 2010. 2011 er dermed det første virkeåret, mens 2019 er siste året vi har helårstall for. Statistikken er avgrenset til fagfellevurderte artikler i vitenskapelige tidsskrift og serier. Tre tema diskuteres særskilt, det er: fordelingen mellom «Gull OA» og «Grønn OA», administrasjonen av UiTs publiseringsfond og vitenarkivet Munin.

Bakgrunn

Open Access har vært viktig for Universitetsbiblioteket ved UiT Norges arktiske universitet i mange år. De første tiltak ble satt i verk allerede før begrepet ble utmyntet tidlig på 2000-tallet. Et forsøk med ETD (Electronic Theses and Dissertations) ble startet rundt år 2000, mens det første OA-tidsskriftet, *Nordlyd*, ble distribuert via UB i 2003. Siden har dette blitt innlemmet i UB-tjenesten Septentrio Academic Publishing, hvor det har fått følge av mange flere tidsskrifter og skriftserier. ETD er på sin side blitt innfaset i tjenesten *Munin*, et vitenarkiv hvor ikke bare masteroppgaver og doktorgrads-avhandlinger, men også vitenskapelige artikler og annet materiale fra UiTs forskere og studenter gjøres åpent tilgjengelig av UB.

I denne oversikten er det likevel naturlig å ta *UiTs første OA-policy* som utgangspunkt. Den ble utarbeidet i fellesskap av UB og Forskningsavdelingen i Sentraladministrasjonen, og ble vedtatt av Universitetsstyret i oktober 2010. Her heter det at UiTs forskere og studenter «som hovedregel» skal velge publiseringskanaler som enten er OA eller som tillater egenarkivering og at de «som hovedregel» skal egenarkivere sine publikasjoner i Munin. Med policyen fikk UB et nytt verktøy og en legitimering av sitt arbeid med OA. Policyen fikk virkning fra og med januar 2011 (se Vedlegg 1 for hele teksten på norsk og engelsk). En policy er ikke en lovtekst eller et bindende dokument i strengt juridisk forstand. Den skal likevel virke som en rettesnor og legge føringer for virksomheten ved statlige institusjoner

https://doi.org/10.7557/15.5506

som UiT. Siden mye handler om opplysningsarbeid og holdningsendringer, er det naturlig at effekten først oppstår over tid.

Septentrio Academic Publishing er ment som en tjeneste for fagmiljøer ved UiT som ønsker å utgi egne skriftserier. UBs rolle her er å gi teknisk og publiseringsfaglig støtte til UiTs forskere i rollen som redaktører. Det er ikke i seg selv et tiltak som gjør store utslag på volumet av åpen publisering ved institusjonen, av to grunner: For det første er det mange av skriftseriene i Septentrio som ikke er fagfellevurderte og dermed ikke telles med i statistikken over vitenskapelig publisering. For det andre er mange av forfatterne til de fagfellevurderte Septentrio-tidsskriftene hjemmehørende ved andre institusjoner; redaktørrollen er utvilsomt et viktig tiltak for å fremme åpen vitenskap, men gir seg ikke utslag i publiseringspoeng ved egen institusjon. Tjenesten Septentrio blir for øvrig beskrevet i en egen artikkel i dette Ravnetrykket og vil ikke behandles videre her.

For å hindre at økonomiske barrierer skulle stå i veien for forskernes publisering i åpne tidsskrifter, ble det fra starten satt av penger til et eget *publiseringsfond*. Fondet skulle dekke såkalte Article Processing Charges (APC), en publiseringsavgift som innkreves av en del åpne tidsskrifter særlig innen natur- og helsefagene. Fondets administrasjon ble lagt til UB. Universitetsstyret besluttet å sette av 300 000 kr til fondet på UiTs budsjett for 2011. Beløpet er senere høynet flere ganger, til dagens forbruk på om lag 4 500 000 kr pr. år. Nøkkeltall i fondets utvikling år for år blir presentert nærmere nedenfor.

Arbeidet med vitenarkivet Munin ble i oppstartsfasen lagt til en egen avdeling ved UB, IFU (Avdeling for IT, formidling og utvikling). Stein Høydalsvik og andre nøkkelpersoner ved IFU fikk etablert en DSpace-installasjon og dirigert personalressurser til å ta unna det som til å begynne med var en relativt begrenset tilflyt av materiale. Imidlertid er det utarbeidet nasjonale og internasjonale standarder for slike institusjonelle vitenarkiv som fordrer utstrakt manuell kontroll med alt som skal tilgjengeliggjøres på nett, så selv et begrenset omfang krevde mye oppbygging av kompetanse. Etter intern omorganisering av UB er det daglige arbeidet med godkjenning og tilgjengeliggjøring av dokumenter i Munin i stor grad overtatt av bibliografisk ekspertise ved Avdeling for samlinger, i tett dialog og samarbeid med Avdeling for publikumstjenester, herunder noen av nøkkelpersonene fra «gamle IFU». Vitenarkivet Munin har også en innleveringsportal for masteroppgaver og doktorgradsavhandlinger ved UiT; her har volumet vært stort i mange år og håndteringen har i stor grad vært ivaretatt av bibliografisk ekspertise ved UB. Det nye er at volumet på andre dokumenttyper enn master- og doktorgradsarbeid har økt markant de siste årene. Dedikert arbeid fra noen få personer over tid har vist seg avgjørende for å sikre kompetanseutvikling og unngå for stort etterslep på arbeidslistene. Omfanget av vitenskapelige artikler i Munin vil bli nærmere presentert nedenfor.

UiTs tall for Grønn og Gull OA

«Grønn Open Access» er OA i regi av forfatteren, det vi gjerne kaller egenarkivering. Ved UiT vil det si at forskeren laster opp (*deponerer*) en versjon av artikkelen – gjerne en manusversjon, det er det de fleste utgivere tillater – i Cristin. Dokumentet blir så eksportert fra Cristin til det åpne vitenarkivet Munin og sjekket nærmere av ansatte ved UB før det tilgjengeliggjøres der.[1]

«Gull OA» er OA hos utgiver. Dette skjer vanligvis i form av publisering i tidsskrift eller andre serier som er OA i sin helhet (heretter kalt OA-kanaler). Noen slike kanaler krever forfatterbetaling (APC); mer om det nedenfor.

«Hybrid OA», hvor enkeltartikler mot betaling blir OA i kanaler som ellers ikke er OA, er en form for Gull OA som ikke støttes av UiTs publiseringsfond. Det kan allikevel skje at enkeltforskere betaler denne formen for OA over eget annuum eller prosjektmidler. Fra og med 2019 har de nye norske konsortieavtalene (Unit-avtalene, jf. Lene Ottesens artikkel), der kostnader for hybrid OA dekkes tilnærmet automatisk, bidratt til at Hybrid OA har vokst.

I oversikten nedenfor representerer tallene for Gull OA kun det som er publisert i rene OA-kanaler. Hybrid OA inngår i tallene for Grønn. Dette skyldes at hybrid så langt ikke har vært mulig å identifisere sikkert i datamaterialet, mens artikler i rene OA-kanaler er enkle å identifisere.

Unit-avtalene innebærer at artikler «kjøpes fri» og blir gull OA på kanalens nettsted, samtidig som egenarkivering tillates. Men at egenarkivering tillates, fører ikke til at artiklene kommer ut i Munin per automatikk; den korresponderende forfatteren må fremdeles laste opp fulltekst i Cristin.

En annen og viktigere forklaring på økningen i Grønn OA er trolig at at UiTs OA-policy ble revidert i 2018 (se Vedlegg 2a for hele teksten på norsk, Vedlegg 2b for hele teksten på engelsk). Revisjonen trådte formelt i kraft med rektors vedtak (på styrets fullmakt) i februar 2019, men opplysningsarbeidet startet flere måneder før den tid. I revisjonen er egenarkivering av tidsskriftartikler nevnt som et særskilt krav. I forbindelse med denne revisjonen har det foregått et nokså intensivt informasjonsarbeid overfor UiTs fagmiljøer om viktigheten av egenarkivering. Temaet er bl.a. tatt opp i UBs kurs rettet mot doktorgradsstipendiater, i faggruppemøter, instituttleder-møter, møter med forskningsgrupper og lignende.

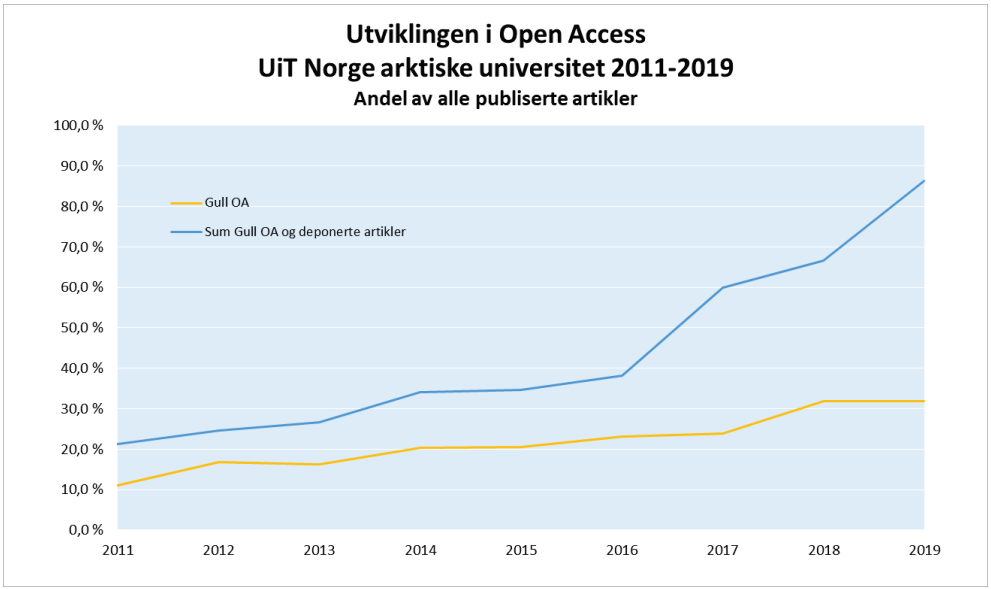
[1] Har artikkelen mer enn én forfatter, holder det at en av forfatterne laster opp. Dette kan være en forfatter ved en annen institusjon som benytter Cristin som verktøy for formidling av artikler til vitenarkiv, slik at institusjonene nyter godt av hverandres innsats på dette området. Det er nå på tale å innføre et nasjonalt vitensarkiv, som muligens vil gjøre etterarbeidet i slike tilfeller enda mindre arbeidskrevende enn det er i dag.

År	2011	2012	2013	2014	2015	2016	2017	2018	2019
Tellende artikler	1194	1285	1284	1361	1533	1637	1663	1819	1868
Åpne og deponerte artikler	253	316	343	463	530	625	998	1212	1613
Andel åpne og deponerte artikler	21,2 %	24,6 %	26,7 %	34,0 %	34,6 %	38,2 %	60,0 %	66,6 %	86,3 %
Gull OA	131	216	209	277	315	379	398	579	594
Andel Gull OA	11,0 %	16,8 %	16,3 %	20,4 %	20,5 %	23,2 %	23,9 %	31,8 %	31,8 %
Deponerte artikler (Grønn OA)	122	100	134	186	215	246	600	639	1019
Andel deponerte artikler	10,2 %	7,8 %	10,4 %	13,7 %	14,0 %	15,0 %	36,1 %	35,1 %	54,6 %
Artikler i TA	1063	1069	1075	1084	1218	1258	1265	1240	1274
Andel deponerte av TA	11,5 %	9,4 %	12,5 %	17,2 %	17,7 %	19,6 %	47,4 %	51,5 %	80,0 %

Figur 1 Tall for Gull og Grønn OA 2011–2019

I Figur 1 er *tellende artikler* totalt antall fagfellevurderte artikler i vitenskapelige tidsskrift og serier (heretter kalt kanaler). *Åpne og deponerte artikler* er summen av artikler som er gjort åpent tilgjengelig gjennom rene OA-kanaler, og artikler publisert i abonnementsbaserte kanaler der forfatteren har lastet opp fulltekst i Cristin. *Andel åpne og deponerte artikler* er dette tallet som prosent av tellende artikler.[2] *Gull OA* er artikler i rene OA-kanaler, *andel Gull OA* dette antallet i prosent av tellende artikler. *Deponerte artikler (Grønn OA)* er deponering av artikler fra abonnementsbaserte kanaler, *Andel deponerte artikler* dette antallet som prosent av tellende artikler. *Artikler i TA* er alle artikler som er publisert i abonnements-baserte kanaler (TA = «toll access»). *Andel deponerte av TA* er andelen av artiklene i abonnementsbaserte kanaler som nå er deponert.

[2] Strengt tatt er det ikke slik at samtlige artikler som blir deponert (lastet opp i Cristin), ender opp med å faktisk bli Grønn OA. Noen ganger har forfatter lastet opp forlagets PDF, og anmodninger fra UB om å sende en manusversjon fører ikke fram. Andre ganger har tidsskriftets utgiver en streng policy som setter juridisk stopper for egenarkivering. Mer om dette nedenfor.



Figur 2 Utviklingen i Open Access ved UiT 2011–2019

Figur 2 gir en grafisk fremstilling av Gull OA og totalen for Gull OA-artikler og deponerte artikler (dvs. «Andel åpne og deponerte artikler» i Figur 1). Det viser en markant vekst de siste årene.

Utviklingen i Gull OA avhenger av finansiering for artikkelbetaling (APC), som ved UiT er løst gjennom publiseringsfondet. Men at finansieringen er på plass, er i seg selv ikke nok for at et rent OA-tidsskrift skal bli valgt. Forskeres oppfatning av slike (ofte nyere) publiseringskanalers kvalitet spiller en stor rolle. I tillegg varierer utviklingen av tjenlige og gode OA-kanaler mellom fagområder. F.eks. har rettsvitenskap gjennomgående ikke hatt relevante OA-kanaler å publisere i før helt nylig. Og OA-kanaler er sterkt underrepresentert på nivå 2 i det norske registeret over vitenskapelige publiseringskanaler, ca. 1/6 sammenlignet med situasjonen for abonnementsbaserte kanaler. Likevel har det vært en jevn vekst i bruken av Gull OA ved UiT, inntil utviklingen flatet ut fra 2018 til 2019. Det kan tenkes at Unit-avtalene har beveget noen forskere til å sende artikkelmanus som ellers ville blitt publisert i en ren Gull OA-kanal, inn til en hybrid OA-kanal isteden. De foreløpige tallene for hybrid (jf. Lene Ottesens artikkel) tilsier i hvert fall en vekst i denne formen for OA, i den korte tiden avtalene har vart. I Unit-avtalene er APC forhåndsbetalt, og forskeren får anledning til å publisere OA uten å måtte gå gjennom søknadsprosedyren hos publiseringsfondet.

Suksessen til Grønn OA (egenarkivering) er avhengig av at det store flertallet av utgivere har policyer som tillater dette – det har jevnt over vært tilfellet hele perioden – og at forskerne kjenner mulighetene og prioriterer å benytte seg av dem. UB har gjennom flere år informert forskerne og kjørt e-post-kampanjer for å få forskerne til å deponere sine artikler (dvs. laste opp fulltekst i Cristin). Økningen fra 2011 til 2016 skyldes i hovedsak dette arbeidet. I 2017 begynte Forskningsrådet med tettere oppfølging av hvorvidt prosjekter de finansierte fulgte opp kravene om OA (obligatorisk egenarkivering), noe som

bidro til en markant økning det året, med en viss vekst også året etter. I kjølvannet av OA-policyens revisjon i februar 2019 har Cristin-ansvarlige ved enhetene – de arbeider mye med disse publikasjonene, og sitter tettere på forskerne – vært mye sterkere engasjert i arbeidet. Samtidig har også noen forskningsdekaner «tatt grep» om sine forskere. Dette har gitt merkbare resultater, egenarkiveringsprosenten for UiT er i 2019 blitt en av de høyeste i landet. Det er neppe realistisk å forvente vesentlig høyere tall for OA i 2020, med mindre det settes i verk ytterligere tiltak i retning av å gjøre eksempelvis egenarkivering obligatorisk for at publikasjonene skal telle i den årlige omfordelingen av midler til enhetene (RBO, resultatbasert omfordeling basert på antall publiseringspoeng). Egenarkivering er i dag et krav fra både Forskningsrådet og EU, men ikke påkrevet i streng juridisk forstand for den forskningen ved UiT som ikke finansieres gjennom eksterne midler.

Publiseringsfondet

UiTs publiseringsfond ble opprettet av UiTs styre i forbindelse med OA-policyen av 2010.

Kriteriene for tildeling er i hovedtrekk (for fullstendig oversikt over fondets reglement, se Vedlegg 3):

- Det må være en ren OA-kanal, fondet dekker ikke avgifter til hybrid-publisering.
- Korresponderende forfatter er søker, og hører til UiT. Både vitenskapelig tilsatte, administrativt tilsatte, ph.d.- og masterstudenter og emerituser kvalifiserer, så lenge publikasjonen gir poeng til UiT.
- Tidsskriftet skal ha nivå 1 eller 2 i det norske registeret over publiseringskanaler, og være registrert i det internasjonale DOAJ (*Directory of Open Access Journals*, en «hvitliste» for seriøse OA-kanaler).
- Fondet dekker i utgangspunktet ikke mer enn 25 000 for en artikkel, dette er en regel som ble innført for noen år siden. Nylig ble det vedtatt at maksgrensen kan fravikes dersom UiT har en formell avtale med utgiver. Denne justeringen ble utløst av Unit-avtalene hvor OA er svært dyrt, men man ikke betaler per artikkel. Det var viktig å gjøre publisering i rene OA-kanaler like enkelt for forfatterne.
- Fondsadministrasjonen kan i velbegrunnede tilfeller gjøre unntak fra enhver regel.

UiTs publiseringsfond ble opprettet i og med at universitetsstyret tildelte 300 000 kr til fondet i budsjettåret 2011. Siden vi ikke hadde noen erfaring med dette, må beløpet kunne sies å være basert på en kombinasjon av røffe anslag og vill gjetting. Fondet har ennå ikke kommet i den situasjon at det har gått tomt for midler og måttet avvise søknader av den grunn, en situasjon vi hører har oppstått på noen andre institusjoner. I 2013 var fondet riktignok i ferd med å ha forpliktet hele årets budsjett på 500 000 allerede i mai, men da grep

rektoratet inn og ga fondet en ekstratildeling på 500 000 slik at budsjettet for 2013 ble 1 000 000, som viste seg å være nok. Erfaringer har vist at det har vært veldig vanskelig å forutse neste års behov, ikke minst når disse behovene skal meldes inn 7–8 måneder i forveien.

Fondet fører et «skyggeregnskap» for å holde oversikt over hvor store beløp som er forpliktet. Forpliktelsene kommer ofte ikke til utbetaling i samme regnskapsår, gjennomgående blir bare 50–60 % av innvilgelsene regnskapsført samme år, de resterende beløp fordeles ut over de følgende 2–3 år med hoveddelen i året etter innvilgelsen. Dette skyldes at noen manus må gå mange runder hos tidsskriftet før de til slutt blir antatt – eller oppgis av forfatteren (jf. Bård Smedsrød og Leif Longvas artikkel).

I tallene nedenfor er alle søknader som formelt kommer inn til fondet og blir registrert i ePhorte, tatt med. Enkle forespørsler som resulterer i at det ikke kommer noen søknad til fondet, blir ikke ført i ePhorte og er heller ikke tatt med i statistikken. For utførlig redegjørelse for detaljene i tabellene, se vedlegg 4.

År	Søknader	Ikke innvilget	Innvilget	Bortfalt	Gjenstående innvilgelser	Bokført	Gjenstående forpliktet
2011	26	2	24	1	23	23	
2012	53	5	48	10	38	38	
2013	101	14	87	18	69	69	
2014	130	4	126	32	94	94	
2015	137	21	116	18	98	98	
2016	209	21	188	50	138	134	4
2017	254	25	229	45	184	164	20
2018	296	19	277	23	254	199	55
2019	297	18	279	38	241	146	95

Figur 3 UiTs publiseringsfond: Antall søknader om dekning av APC, 2011–2019

År	Budsjett	Ref NFR	Total ramme	Søknader	Ikke innvilget	Innvilget	Bortfalt	Gjenstående innvilgelser	Bokført	Gjenstående forpliktet	Resultat	Akkumulert resultat
2011	300		300	203	32	172	14	157	157	0	143	
2012	309		309	491	61	430	104	326	326	0	-17	126
2013	1 000		1 000	1 187	194	993	256	737	737	0	263	246
2014	1 300	491	1 791	1 809	50	1 759	525	1 233	1 233	0	558	820
2015	1 300	666	1 966	2 198	423	1 776	322	1 453	1 453	0	513	1 070
2016	1 400	763	2 163	3 802	564	3 238	985	2 253	2 205	48	-90	423
2017	1 400	1 033	2 433	4 263	620	3 643	779	2 864	2 534	330	-431	-521
2018	3 500	1 284	4 784	5 379	491	4 888	419	4 469	3 475	994	315	-116
2019	6 300	0	6 300	6 114	490	5 624	941	4 682	2 597	2 085	1 618	1 933

Figur 4 UiTs publiseringsfond: Søknadsbeløp og innvilgelser 2011–2019. Tall i hele 1000 kr.

Vitenarkivet Munin

Som nevnt ovenfor er definisjonen på Grønn OA at det skjer i regi av forfatteren. Det er formelt riktig, uten deponering via Cristin blir ikke artikkelen åpent tilgjengelig i vitenarkivet Munin. Men egenarkivering forutsetter mer enn deponering. Først når artikkelen har gått hele veien fra opplasting (deponering) til tilgjengeliggjøring i Munin, er den reelt blitt egenarkivert og tilgjengelig som en «Grønn OA-artikkel».

Tilgjengeliggjøring av deponerte artikler krever også et arbeid fra andre mennesker enn bare forfatteren. Universitetsbiblioteket som drifter institusjonens vitenarkiv, i vårt tilfelle Munin, må sjekke at lisensen er i orden og eventuelt legge på en embargo (en del utgivere krever 6–12, endog opp til 24 måneder fra artikkelen er publisert til den kan tilgjengeliggjøres i vitenarkivet). Det er også krav om et visst minimum av kontrollerte metadata, for å sikre god gjenfinnbarhet og motvirke usikkerhet om hva det er slags tekst leseren har med å gjøre.[3] Dette arbeidet er ressurskrevende, men langt fra statisk.

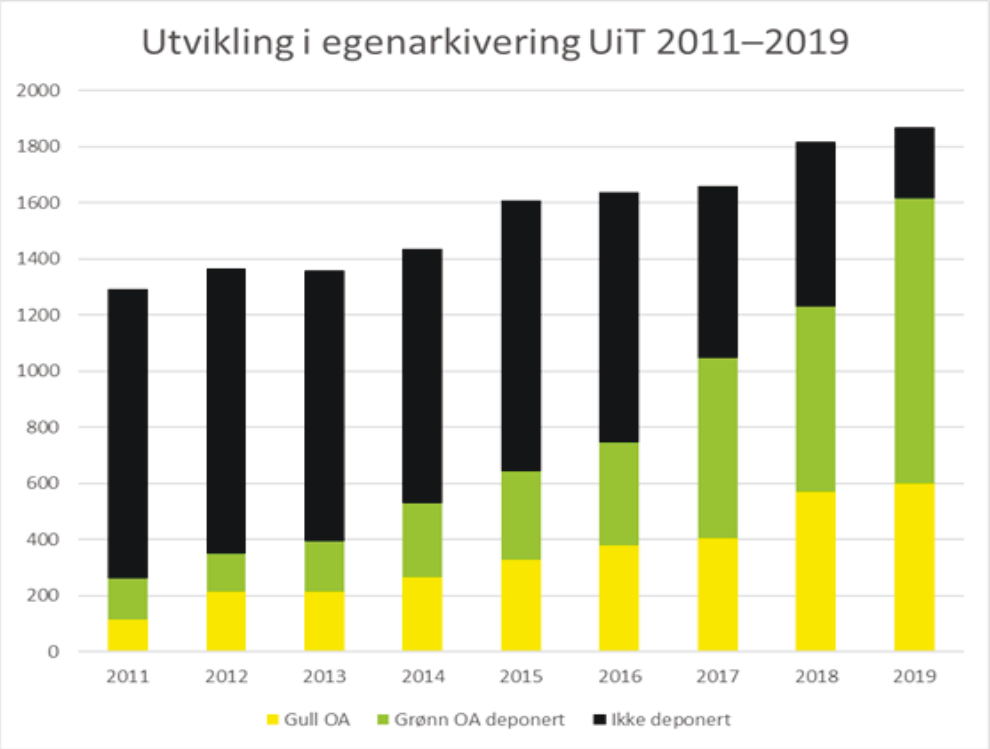
Blant oppgraderinger som har skjedd i senere år er innføring av «Altmetrics-smultring», et verktøy for å måle henvisinger til vitenskapelige arbeider i sosiale medier og andre uformelle kanaler. Det er også gjort flere forbedringer på metadata-siden som vi ikke skal gå gjennom her. Et interessant spørsmål er imidlertid om hvor mye av det som lastes opp i Cristin, som faktisk kommer ut i arkivet. Artikler kan bli avvist fordi utgiveren ikke tillater noen form for tilgjengeliggjøring, eller det kan bli avvist fordi forfatteren har lastet opp feil

[3] Ofte dreier det seg om forfatterens siste, fagfellevurderte manusrevisjon. Selv om ordlyden er identisk, har artikkelen en annen layout enn den offisielle versjonen som befinner seg bak betalingsmur på utgiverens hjemmeside.

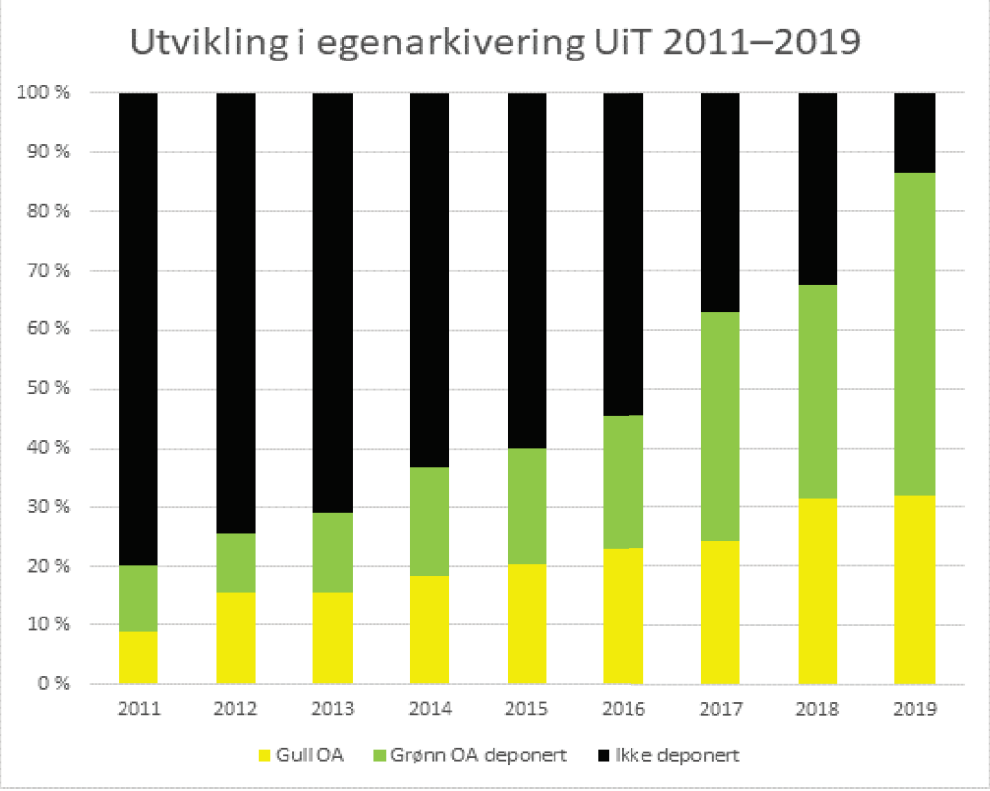
versjon, og på tross av henvendelser fra bibliotekets side ikke leverer den manusversjonen som etterspørres.

Det skal innrømmes at kapasitetsproblemer ved biblioteket fører til at det noen ganger kan gå lang tid fra deponering til tilgjengeliggjøring. Her skjer det en tett dialog mellom Munin-teamet og bibliotekets ledelse for å vurdere tiltak. Munin inneholder mange dokumenttyper, der fagfellevurderte, vitenskapelige artikler bare er en av flere. De prioriteres høyt, høyere enn konferansebidrag, kronikker, rapporter og annet «grått» materiale, men ikke like høyt som doktorgradsavhandlinger og mastergradsoppgaver, som behandles umiddelbart etter at studieadministrasjonen ved aktuelt institutt/fakultet har ferdigbehandlet oppgaven. Kontakt med forfattere som har lastet opp en annen versjon enn den utgiver tillater, medfører ofte også noe forsinkelse. Til tross for tiden som går med til dette, har UB ved UiT som politikk å alltid prøve i det lengste. Dette ut fra en erkjennelse av at faktisk tilgjengeliggjøring er en prioritert oppgave for UB og et strategisk mål for UiT. Den desidert viktigste årsaken til at det kan gå lang tid fra deponering til tilgjengeliggjøring, er imidlertid de embargoperiodene som forlaget setter. Slike artikler blir ferdigbehandlet fra Munin-teamets side, men blir da liggende i arkivet med en «tikkende hengelås» som åpnes automatisk den dagen embargoperioden er over.

I tabellene nedenfor vises utviklingen for perioden 2011–2019. Gul farge markerer artikler i rene OA-kanaler (*Gull OA*). *Grønn OA deponert* vil si artikler som er blitt lastet opp; en svært liten andel av disse vil ikke komme ut på grunn av problemer med rettigheter eller manglende manusversjon. Markert med svart er artikler som er publisert i abonnementsbaserte kanaler, og forfatteren har unnlatt å laste opp fulltekst i Cristin (*ikke deponert*).



Figur 5 Utvikling i egenarkivering av fagfellevurderte artikler skrevet av forskere tilknyttet UiT, perioden 2011–2019 (absolutte tall)



Figur 6 Utvikling i egenarkivering av fagfellevurderte artikler skrevet av forskere tilknyttet UiT, perioden 2011–2019 (prosentvis fordeling)

Veien videre

I sum er OA-arbeidet ved UiT profesjonelt ivaretatt. Grepet med etablering av et eget publiseringsfond som aldri har gått tomt, har gjort at UiTs forskere slipper å belaste eget annuum for å publisere i rene OA-kanaler som krever APC. Tett kontakt med enkeltforskere og systematisk arbeid med å holde etterslepet på arbeidslistene i Munin-tjenesten nede er et annet viktig grep.

Informasjonsarbeidet har forbedringspotensial, skjønt det skal sies at det ikke bare er Universitetsbiblioteket som har ansvar for denne virksomheten. I den sammenhengen er det gledelig å konstatere at dialogen med Cristin-ansvarlige og forskningsdekaner ved fakultetene er blitt stadig bedre, ikke minst i kjølvannet av revisjonen av OA-policyen og innføringen av Unit-avtalene i 2019. I 2020 har vi også lansert en egen portal for publisering og Open Access ved UiT. Effekten av dette tiltaket skal det bli spennende å følge med på.

Mengden artikler som lastes opp for tilgjengeliggjøring i Munin har økt år for år. Samfunnsøkonomisk vil nok det mest rasjonelle være at andelen artikler som publiseres i rene OA-kanaler beveger seg oppover fra det foreløpige toppnivået på knappe 32 %. Enn så lenge er det likevel gledelig at stadig flere deponerer sine artikler og bidrar til at ny kunnskap blir åpent tilgjengelig som grønn OA. Det er likevel et stykke igjen til 100 % åpen tilgjengeliggjøring av vitenskapelige artikler som skrives ved UiT. Her kan institusjonen vurdere å vedta både «pisk» og «gulrot».

Insentivordninger hvor institutter og forskningsgrupper som kommer best ut på statistikken, premieres økonomisk og gis publisitet, kan virke positivt. Å holde tilbake publiseringspoeng for artikler som verken publiseres i en OA-kanal eller egenarkiveres, er et annet tiltak som kan gi gode resultater raskt. Få forskere vil nok ønske å straffe sitt eget institutt økonomisk ved å ikke utføre den lite tidkrevende jobben det er å laste opp fulltekst i Cristin.

En annen mulighet er å tenke at kritisk masse allerede er oppnådd. Toppåret 2019, hvor 86 % av de vitenskapelige artiklene ved UiT enten kom ut i rene OA-kanaler eller ble lastet opp i fulltekst (deponert), viser at publiseringskulturen ved institusjonen jevnt over er OA-vennlig. Kulturendringen har skjedd raskt, dette tallet lå i 2011 på 20 %. Sammen med utviklingen nasjonalt og internasjonalt, med Unit-avtaler og Plan S som viktige drivende faktorer, er vår forskning på god vei mot en virkelighet der OA blir standarden. Men det er ennå et stykke igjen til regjeringens mål om at alle artikler som skrives av forskere tilknyttet norske institusjoner, skal være åpent tilgjengelige innen 2024. Vi er heller ikke i mål med UiTs strategi fram mot 2022, Drivkraft i nord, hvor det heter at «UiT skal være nasjonalt ledende på Open Science og våre forskningsdata og publikasjoner skal være åpent tilgjengelig der dette er mulig». Skal dette målet nås, må trykket holdes oppe og en ytterligere revisjon av UiTs OA-policy vurderes.

Til sist vil vi understreke at arbeidet med OA ved UB hele veien har nytt godt av sterk støtte fra universitetets og bibliotekets ledelse. Verken policydokumenter eller strategier vokser ut av noe vakuum, men fordrer tett dialog og samhandling mellom de ulike nivåene i en organisasjon. Stein Høydalsvik har hele veien spilt en uvurderlig rolle i det strategiske og faglige arbeidet med hele sakskomplekset åpen vitenskap. Bibliotekdirektørene Helge Salvesen og Johanne Raade har vært tydelige og handlekraftige støttespillere. Og ikke minst har rektoratene under både Jarle Aarbakke, med Curt Rice som prorektor for forskning, og Anne Husebekk, med Kenneth Ruud som prorektor, gitt politisk drahjelp både internt ved institusjonen og eksternt.

Ravnetrykk 39, 2020

Universitetet i Tromsøs prinsipper for Open Access publisering

Vedtatt av:	Universitetsstyret i sak S 33-10
Gjelder fra:	14. oktober 2010
Arkivreferanse:	2010/5305-1

Fri tilgang til vitenskapelige resultater er en viktig forutsetning for et velfungerende demokrati, for fri meningsutveksling og for at vitenskapen skal kunne være et godt verktøy for utvikling av samfunns- og næringsliv. UiT har som mål at alle vitenskapelige publikasjoner fra universitetet skal gjøres tilgjengelig enten i Open Access-tidsskrifter eller i åpne vitenarkiver. Universitetet i Tromsø slutter seg derfor til følgende grunnleggende Open Access-prinsipper om fri tilgang til vitenskapelige resultater.

- **Egenarkivering:** UiT har som hovedregel at studenter og forskere egenarkiverer sine publikasjoner i Munin. Tilgjengeliggjøring i Munin vil skje innenfor rammen av de avtaler som forfatteren har inngått med publikasjonens utgiver og utgivernes prinsipper for egenarkivering. Universitetsbiblioteket har ansvar for å avklare rettighetsspørsmål knyttet til tilgjengeliggjøring.

- **Valg av publiseringskanal**: UiT har som hovedregel at studenter og forskere skal, under forutsetning av faglig likeverdighet, velge de publiseringskanaler som gir friest mulig tilgang til publikasjonene, enten ved at de har en god politikk med hensyn til å tillate egenarkivering eller ved at publiseringskanalen er en Open Access publiseringskanal.

- **UiT som utgiver**: Universitetet skal arbeide for at tidsskrifter og skriftserier som utgis ved UiT skal være Open Access-utgivelser. Målet er at alle tidsskrifter og skriftserier som utgis ved UiT skal tillate egenarkivering, og det skal oppmuntres til at det er den endelig utgitte utgaven (forleggers PDF) som arkiveres.

University of Tromsø principles for Open Access publishing[1]

Free access to scientific results is an important prerequisite for a well-functioning democracy, for the free exchange of opinions and to enable science to be a tool for the development of society and industry. The University of Tromsø has as its goal that all scientific publications from the university shall be made available either in an Open Access journal or in an institutional repository. Hence, the University of Tromsø will adhere to the following basic principles of Open Access to scientific results.

- **Self-archiving:** The University of Tromsø has as its general rule that students and researches shall self-archive their publications in Munin, the university's institutional repository. Publications will be made available through Munin within the constraints of the agreements the authors have made with the publisher and publishers' principles for self-archiving. The University Library of Tromsø has the responsibility for investigating and ensuring compliance with publishers' policies and other questions regarding intellectual property rights in this context.

- **Choice of publishing venue**: The University of Tromsø has as its general rule that students and researchers shall – provided publications are of equal scientific worth and stature – choose publishing venues that provide the freest access to the publications, either through having a positive attitude to self-archiving or by being an Open Access publishing venue.

- **The University as a publisher:** The University shall endeavour to make journals and other publications published by the University Open Access publications. It is a goal that all publications published by the University shall permit self-archiving, and that self-archiving of the final published version (publisher's PDF) shall be encouraged.

[1] This is a translation of the document approved by the Board of the University of Tromsø on October 14th 2010 (ref. S 33-10); this English language version has not been formally approved.

UiT
NORGES
ARKTISKE
UNIVERSITET

PRINSIPPER FOR ÅPEN TILGANG TIL VITENSKAPELIGE PUBLIKASJONER VED UIT NORGES ARKTISKE UNIVERSITET

Fastsatt av:	Universitetsstyret	Dato:	14.10.2010 (S 33-10), 07.02.2019 (F 03-19)
Ansvarlig enhet:	Avdeling for forskning, utdanning og formidling	Arkivref.:	2018/5399
Sist endret av:	Avdeling for forskning, utdanning og formidling	Dato:	07.02.2019

Fri tilgang til vitenskapelige resultater er en viktig forutsetning for et velfungerende demokrati, for fri meningsutveksling og for at vitenskapen skal kunne være et godt verktøy for utvikling av samfunns- og næringsliv. UiT Norges arktiske universitet har en sterk nasjonal posisjon som fremmer av åpenhet til forskningsresultater, både publikasjoner og forskningsdata, og universitetsledelsen har en viktig rolle som nasjonal og internasjonal pådriver for å nå mål om åpen tilgang. Alle vitenskapelige publikasjoner fra UiT skal gjøres tilgjengelige enten i publiseringskanaler med åpen tilgang eller i åpne vitenarkiver. UiT har derfor nedfelt følgende prinsipper om fri tilgang til vitenskapelige publikasjoner:

1. Valg av publiseringskanal

UiTs ansatte og studenter medvirker i vitenskapelig publisering i ulike roller, – som forfatter, fagfelle, medlem av redaksjonsråd eller redaksjonsmedlem. UiT ønsker at man prioriterer deltakelse i publiseringskanaler som fremmer åpen tilgang fremfor hos utgivere som ikke fremmer dette prinsippet.

UiTs ansatte og studenter skal publisere i kanaler som gir friest mulig tilgang til publikasjonene. Publikasjonskanalen skal enten gi åpen tilgang på egen plattform (såkalt gull åpen tilgang) eller ha en god politikk med hensyn til egenarkivering (såkalt grønn åpen tilgang). Åpne publiseringskanaler skal velges fremfor lukkede publiseringskanaler under forutsetning av faglig forsvarlighet. Dette er et krav for artikler i fagfellevurderte tidsskrifter, men anbefales også for vitenskapelige bøker (monografier og antologier), samt for populærvitenskapelige artikler, aviskronikker og annen formidlingsvirksomhet.

UiT bidrar til å dekke eventuelle ekstrakostnader til åpen publisering gjennom et eget publiseringsfond, under forutsetning av at publiseringsfondets vilkår oppfylles.

2. Egenarkivering

UiTs ansatte skal deponere (laste opp) sine publikasjoner fortløpende i Cristin, slik at de om mulig kan tilgjengeliggjøres i UiTs vitenarkiv Munin. Dette kravet gjelder foreløpig kun artikler i fagfellevurderte tidsskrifter, men anbefales også for vitenskapelige monografier og antologiartikler. Universitetsbiblioteket har ansvar for å avklare rettighetsspørsmål knyttet til tilgjengeliggjøring av publikasjoner i Munin. Tilgjengeliggjøring i Munin vil skje innenfor rammene av utgivers generelle prinsipper for egenarkivering, med mindre den ansatte har inngått

særlige avtaler for sin publikasjon. På sikt vil andelen deponerte og tilgjengeliggjorte artikler kunne få betydning ved UiTs fordeling av resultatbaserte budsjettmidler til fakultetene, i tråd med Regjeringens mål og retningslinjer[1].

Masteroppgaver og doktoravhandlinger leveres inn til bedømming gjennom innleveringsportalen i Munin. Universitetsbiblioteket sørger for å gjøre disse åpent tilgjengelige, med mindre det er gjort særlig avtale om unntak fra tilgjengeliggjøring.

3. Evaluering

Ved evalueringer i forbindelse med tilsetting, opprykk eller ved tildeling av doktorgrad skal det legges vekt på vitenskapelige arbeiders faglige innhold med hensyn til kvalitet, relevans og betydning, i henhold til prinsippene i The San Francisco Declaration on Research Assessment[2].

Ved andre evalueringer og ved tildelinger av forskningsmidler fra UiT, skal de samme prinsipper legges til grunn. Ved tildeling av FoU-termin skal søkere oppfylle krav til vitenskapelig publisering. Dersom det er behov for å prioritere mellom søkere, er andel publikasjoner som er publisert i åpne publiseringskanaler eller som er deponert for tilgjengeliggjøring i vitenarkiv blant kriteriene som skal tillegges vekt.

4. UiT som utgiver og innkjøper

Innen utgangen av 2021 skal alle vitenskapelige publiseringskanaler der UiT er utgiver eller medutgiver, ha åpen tilgang i den forstand at de minst skal tillate umiddelbar egenarkivering av publisert versjon (med mindre annet er nødvendig av juridiske grunner). Tilsvarende skal vitenskapelige tidsskrifter og bokserier hvor UiT står som eier eller medeier, innen utgangen av 2021 være konvertert til utgivelser som enten selv gir, eller tillater, åpen tilgang til publisert versjon.

Universitetsbiblioteket skal ved inngåelse av avtaler med utgivere om elektroniske abonnementer og lignende, alene eller gjennom samarbeid med andre institusjoner, etterstrebe å sikre gode betingelser for åpen tilgang. Gode betingelser innebærer bærekraftige priser i åpne tidsskrifter, eller åpen tilgang uten ekstra kostnad i abonnementstidsskrifter, eventuelt i form av tjenlige vilkår for egenarkivering i Munin. Denne typen avtaler skal ikke kunne unntas offentlighet.

---

[1] Nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler, av 22.08.2017 https://www.regjeringen.no/no/dokumenter/nasjonale-mal-og-retningslinjer-for-apen-tilgang-til-vitenskapelige-artikler/id2567591/

[2] The San Francisco Declaration on Research Assessment (DORA-erklæringen), 2012 https://sfdora.org/read/

UiT
NORGES
ARKTISKE
UNIVERSITET

PRINCIPLES FOR OPEN ACCESS TO ACADEMIC PUBLICATIONS AT UIT THE ARCTIC UNIVERSITY OF NORWAY

Approved by:	The University Board	Date:	14.10.2010 (S 33-10), 07.02.2019 (F 03-19)
Unit responsible:	Research, Education and Communication Division	Archive ref.:	2018/5399
Last amended by:	Research, Education and Communication Division	Date:	07.02.2019

Open access to research results is an important prerequisite for a well-functioning democracy, for the free exchange of opinions and to enable the research to be a tool for the development of society and industry. UiT The Arctic University of Norway has a strong position nationally as a promoter of open access to research results, publications as well as research data, and the university management has an important role as a national and international driving force for achieving the goal of open access. All academic publications from UiT shall be accessible in open access journals or in institutional repositories. Consequently, UiT has established the following principles concerning free access to academic publications:

1. Choice of publishing channel

UiT's staff and students have various roles in academic publishing, including as author, peer, member of an editorial board or editorial member. UiT wants its staff and students to prioritise publishing channels that promote open access over those that do not promote this principle.

UiT's staff and students shall choose publishing channels that provide the highest level of open access. The publishing channel shall either provide open access on its own platform (gold open access) or have a satisfactory policy concerning self-archiving (green open access). Open publishing channels shall be chosen in preference to closed subscription-based publishing channels where academically acceptable. This is a requirement for articles published in peer-reviewed journals, while it is also recommended for academic books (monographs and anthologies), as well as popular science articles, feature articles in newspapers and other dissemination activities.

UiT contributes to covering any additional costs associated with open access publishing through its own publishing fund, providing the publishing fund's conditions are fulfilled.

2. Self-archiving

UiT's staff shall continuously deposit (upload) their publications in Cristin so they may be accessed in UiT's institutional repository. This requirement currently applies only to articles published in peer-reviewed journals, while it is also recommended for academic monographs and anthologies. The University Library is responsible for clarifying matters regarding to intellectual

property rights associated with publishing in Munin. Access in Munin will occur within the scope of the publishers' general principles for self-archiving, unless staff members have entered into special agreements concerning their publication. In the long term, the proportion of deposited and accessible articles may be of significance for UiT's allocation of results-based budgets to the faculties, in line with the Government's goals and guidelines[1].

Master's and doctoral theses are submitted for assessment through the submission portal in Munin. The University Library takes care of making these openly accessible, unless a special agreement has been entered into exempting the thesis from being made openly accessible.

3. Assessments

When making assessments concerning appointments, promotions or the awarding of doctorates, emphasis shall be attached to the quality, relevance and importance of the content of scholarly works, in accordance with the principles of the San Francisco Declaration on Research Assessment.[2]

For other assessments and when allocating research funding from UiT, the same principles shall be used as a basis. When allocating R&D sabbaticals, applicants shall fulfil requirements concerning academic publishing. Where the need arises to rank applicants, emphasis shall be attached to criteria such as the proportion of publications that are published in open publishing channels or deposited for access in open repositories.

4. UiT as publisher and purchaser

By the end of 2021, all academic publishing channels of which UiT is the publisher or co-publisher shall provide open access in the sense that, as a minimum requirement, they must allow immediate self-archiving of the published version (unless otherwise necessary for legal reasons). Similarly, by the end of 2021, all scientific journals and book series owned or co-owned by UiT shall be converted into publications that either provide or permit open access to the published version.

When entering into agreements with publishers concerning electronic subscriptions, etc., either alone or through cooperation with other institutions, the University Library shall strive to ensure the best possible conditions for open access. In practice, this means sustainable prices in open journals, or open access at no extra cost in subscription-based journals, possibly in the form of suitable conditions for self-archiving in Munin. This type of agreement must not be exempt from public disclosure.

---

[1] National goals and guidelines for open access to research articles, of 22.08.2017 https://www.regjeringen.no/no/dokumenter/nasjonale-mal-og-retningslinjer-for-apen-tilgang-til-vitenskapelige-artikler/id2567591/

[2] The San Francisco Declaration on Research Assessment (DORA), 2012 https://sfdora.org/read/

UiT
NORGES
ARKTISKE
UNIVERSITET

Universitetsbiblioteket/Avdeling for IT, formidling og utvikling
Arkivref.: 2011/1088
Dato: 10.12.2015

Publiseringsfondet ved UiT – retningslinjer

1. Det kan gis støtte til dekning av utgifter til forfatterbetaling (Article Processing Charges, APC, Article Submisson Charges og lignende) i Open Access-tidsskrifter. I utgangspunktet gis støtten slik at fondet dekker hele utgiften for artikler som innvilges støtte, men dersom APC-beløpet ved innvilgelsen beregnes til å bli mer enn NOK 25 000 inkl. mva. må forfatteren besørge dekning av overskytende beløp og før innvilgelse av støtte meddele fondet hvor restbeløpet skal belastes.

   1.1. Det gis ikke støtte til dekning av utgifter til fargetrykk, ekstra illustrasjoner, stort omfang og lignende.

   1.2. Det gis ikke støtte til publisering i hybridtidsskrifter. Hybridtidsskrifter er tidsskrifter hvor man betaler for å gjøre en enkeltartikkel Open Access, samtidig som tidsskriftet er abonnementsfinansiert.

   1.3. Det kan etter vurdering gis støtte til publisering av artikler/kapitler i antologier/bøker som publiseres slik at publikasjonen er tellende. Det er en forutsetning at hele antologien/boken er Open Access og gratis tilgjengelig. Beløpsgrensen i pkt. 1 gjelder for hvert enkelt bidrag/kapittel, og det kan settes en lavere beløpsgrense etter vurdering.

   1.4. Det kan også gis støtte til OA-publikasjoner/OA-publikasjonsformer som bruker andre finansieringsmodeller enn APC, f.eks. institusjons- eller forfattermedlemskap. Dette vurderes av UB, og rapporteres særskilt om.

   1.5. Det kan gis støtte til lagring av data i tilknytning til publisering av artikler som har fått støtte fra fondet.

   1.6. Det kan etter vurdering gis støtte til publisering av OA-monografier. Maksimumsbeløpet i pkt. 1 gjelder ikke for monografier.

2. Forutsetninger for tildeling

   2.1. Korresponderende forfatter må ha tilhørighet ved UiT gjennom tilsetting eller som student. (Se likevel pkt. 4 nedenfor.) Det stilles ikke krav til stillingsomfanget, men tilhørigheten må også fremkomme i artikkelens angivelse av forfatterens institusjonstilhørighet.

   2.2. Eksternt finansierte prosjekter bør vanligvis også finansiere forfatterbetaling for artikler som er et resultat av prosjektet, og midler til å dekke slike utgifter bør derfor innarbeides i prosjektsøknadene.

Norges forskningsråd bidrar til finansieringen av all artikkelbetaling gjennom fondet. Artikler bygget på forskning finansiert av forskningsrådet skal derfor i denne sammenhengen likebehandles med internt finansiert forskning. All artikkelbetaling for artikler knyttet til forskningsrådfinansiert forskning skal skje gjennom fondet, og denne typen utgifter skal ikke belastes prosjektene.

2.3. Artikkelen må registreres i Cristin, og den publiserte fullteksten lastes opp i Cristin senest ved publiseringstidspunktet.

2.4. Tidsskriftet må være registrert i Directory of Open Access Journals (DOAJ, www.doaj.org) og må være på nivå 1 eller 2 i NSDs register over publiseringskanaler, se (http://dbh.nsd.uib.no/kanaler/). Om tidsskriftet ikke er vurdert av NSD, må forfatteren dokumentere at tidsskriftet er innmeldt for vurdering (http://dbh.nsd.uib.no/kanaler/forslag.do), og at det har rutiner for fagfellevurdering slik at godkjenning på nivå 1 er sannsynliggjort.

3. Det søkes om støtte på skjema som finnes på Universitetsbibliotekets nettsider. Det er ingen søknadsfrist, men det normale vil være at man søker før man sender manus inn til tidsskriftet for vurdering. Universitetsbiblioteket avgjør søknadene løpende og gir tilbakemelding til forfatteren, og informerer om hvordan betaling skal skje.

3.1. Dersom støtte avslås fordi fondets midler er oppbrukt, kan det ikke søkes om dekning av utgifter for den samme artikkelen i et senere år der fondet ev. har fått nye midler.

4. Universitetsbiblioteket kan i velbegrunnede enkelttilfeller innvilge støtte selv om ikke alle krav er oppfylt.

5. Retningslinjene

5.1. Retningslinjene vedtas av Universitetsdirektøren.

5.2. Retningslinjene og praktiseringen av dem vil løpende bli evaluert i samarbeid med bibliotekdirektørens ressursgruppe for publisering.

5.3. Disse retningslinjene vil kunne endres uten varsel, men uten at allerede meddelte innvilgelser eller avslag påvirkes.

5.4. Universitetsbibliotekets avgjørelser er endelige og kan ikke påklages.

Retningslinjene er fastsatt av Universitetsdirektøren. Revidert utgave per 10.12.2015.

VEDLEGG 4: Tillegg til Jan Erik og Per Pippin sitt bidrag om OA UiT 2011–2019

Etter at UNIT-avtalene kom er ikke formelle søknader som viser seg å ha med tidsskrift som er dekket av UNIT-avtalene å gjøre, tatt med, de medfører ikke noen saksbehandling, kun informasjon og veiledning overfor forfatterne.

Over tid har vi følgende kategorier som kommer til syne i tabellene

- Søknader: Alle søknader til fondet som kommer langt nok til å bli ført i ePhorte.
- Ikke innvilgede søknader: Søknader som blir avvist ved første behandling, oftest fordi tidsskriftet er et abonnementstidsskrift (hybrid) eller av andre grunner ikke tilfredsstiller kravene.
- Bortfalte søknader: Søknader som blir innvilget, men som faller bort, som regel fordi manuset ikke blir antatt i det aktuelle tidsskriftet eller hos den aktuelle utgiver. Det forekommer også at forfatteren slipper å betale av ulike grunner, eller at det viser seg at noen andre (eksempelvis forskningsprosjektets driftsmidler) har betalt. Vi legger ikke veldig mye energi i å hindre andre i å betale, selv om vi griper inn om vi oppdager dette innen rimelig tid.
- Gjenstående innvilgelser er søknader som ikke blir avvist og som heller ikke har bortfalt. De fordeler seg i hhv. betalte innvilgelser (der hvor vi har fått faktura) og i gjenstående forpliktelser, dvs. at vi har en forpliktelse hvor vi ennå ikke er fakturert. Alle innvilgelser i perioden 2011–2015 er betalt, for 2016 gjenstår i øyeblikket 4 innvilgelser mens tallene for 2017–2019 er høyere, naturlig nok høyest for 2019. Periodisk gjennomgås eldre gjenstående innvilgelser, og vi utber oss en prognose fra forfatteren for å klarlegge om innvilgelsen bør slettes. Alle innvilgelser har en forfallsdato, men vi sletter ikke innvilgelser om det fortsatt synes å være liv i manuset, det tjener ingen hensikt.

I tillegg til bevilgning fra universitetsstyret, får fondet også tilførsel av midler via Forskningsrådets STIM-OA-ordning. Dette er en ordning hvor UH-institusjoner som har et publiseringsfond kan søke om å få refundert inntil 50 % av regnskapsførte kostnader for foregående regnskapsår. UiT har gjennomgående fått mindre enn 50 %, av to grunner:

- 1. STIM-OA-ordningen har ikke hatt tilstrekkelig med midler, og har avkortet forholdsmessig.
- 2. Vi innvilger støtte til publikasjoner som ikke tilfredsstiller STIM-OA-ordningens krav. Dette gjelder monografier og antologi-artikler (STIM-OA gjelder kun for tidsskriftartikler). Vi innvilger også støtte til artikler i tidsskrift som ikke er godkjent hos NSD og/eller DOAJ, dersom vi får dokumentert at det er søkt om godkjenning, og vi vurderer at tidsskriftet tilfredsstiller kravene. Som regel går dette i orden før vi søker, men i noen tilfeller er tidsskriftene for nye til at NSD vil godkjenne dem og vår refusjonssøknad kommer da ikke inn under STIM-OA-kriteriene.

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Fra abonnement til åpen tilgang

Et blikk på Units konsortieavtaler med krydder fra UiT

Lene Ottesen

Sammendrag

Per mars 2020 har Norge 11 avtaler med akademiske forlag som sikrer umiddelbar åpen tilgang. UiT har sluttet seg til alle disse avtalene. I den spede begynnelse viser publiseringstallene at mer blir åpent tilgjengelig. Tidlige internasjonale erfaringer har pekt på det samme, men også at det koster. Avtalene oppfyller i store trekk nasjonale politiske krav og kravene i Plan S. Likevel er det usikkert hvilken retning vi beveger oss i. Avtaler med forlagene som inkluderer åpen tilgang bidrar til økt åpen publisering, men hvorvidt de også vil bidra til å redusere kostnader på sikt og å endre systemet gjenstår å se.

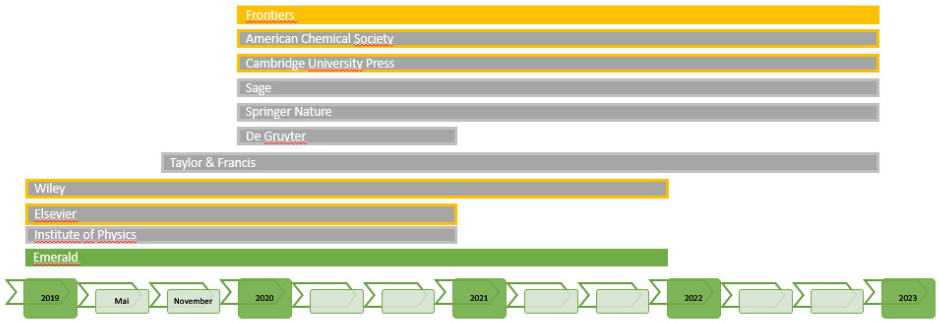
Norge får sine første forlagsavtaler med åpen publisering

Etter at den norske regjering lanserte nasjonale retningslinjer for åpen tilgang til forskningsresultater i august 2017, har bibliotek-konsortiet *Unit lisenser og åpen tilgang* inngått en rekke avtaler med akademiske forlag som inkluderer åpen publisering. Retningslinjene slo fast at

> «Institusjoner og konsortier som forhandler avtaler med forlag skal sørge for at avtalene fremmer åpen tilgang uten økte totalkostnader, og at det er åpenhet om avtalenes betingelser» (Kunnskapsdepartementet, 2017).

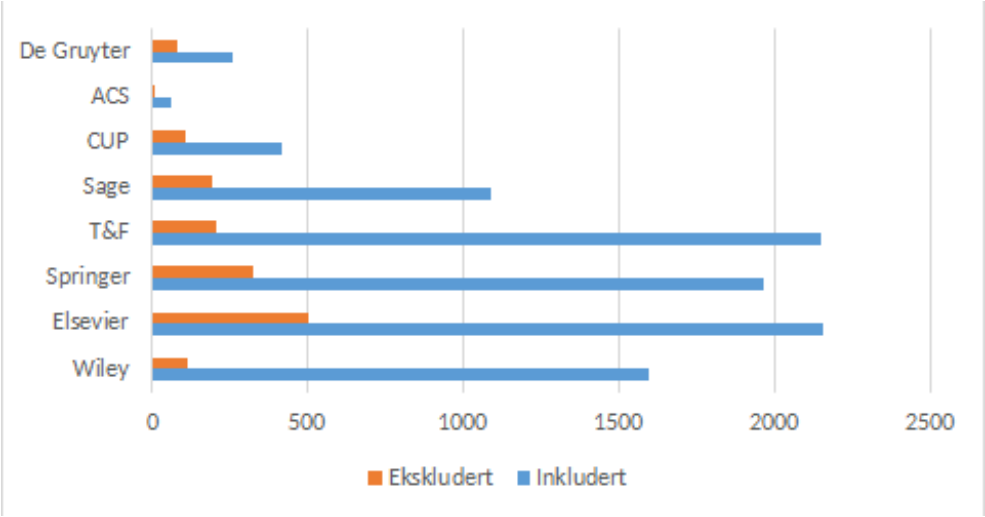
De fleste norske avtalene er såkalte les-og-publiser-avtaler som gir norske forskere både mulighet til å lese abonnementsinnhold og publisere sine artikler åpent. I tillegg er det én avtale som dekker deponering i vitenarkiv uten embargo, samt én ren publiseringsavtale med et åpent forlag. Avtaler som skal sørge for overgang fra abonnement til publisering kalles gjerne transformative avtaler. UiT Norges arktiske universitet har sluttet seg til samtlige av disse nasjonale avtalene.

https://doi.org/10.7557/15.5498



Figur 1: Norske nasjonale avtaler som inkluderer lesetilgang og publisering per mars 2020. Grønn = deponering i vitenarkiv uten embargo. Grå = hybridpublisering. Gull = publisering i åpne tidsskrift.

Avtalene er i stor grad bygget opp på samme måte, selv om de også har sine særtrekk og variasjoner. Detaljert beskrivelse av hovedtrekkene i de ulike avtalene finnes på nettstedet OpenAccess.no. Et gjennomgående trekk som er verdt å merke seg er at avtalene sjelden inkluderer alle tidsskriftene som utgis av forlaget. Noen utelater sine mest eksklusive tidsskrift, tidsskrift eid av vitenskapelige foreninger (*societies*) eller de åpne tidsskriftene. Hvilke tidsskrifter som inngår er gjenstand for forhandling, ikke bare mellom partene, men også mellom forlagene og vitenskapelige foreninger.



Figur 2: Antall tidsskrift fra forlagene som er inkludert og ekskludert i avtalene.

Avtaler med forlagene som en vei til åpen tilgang

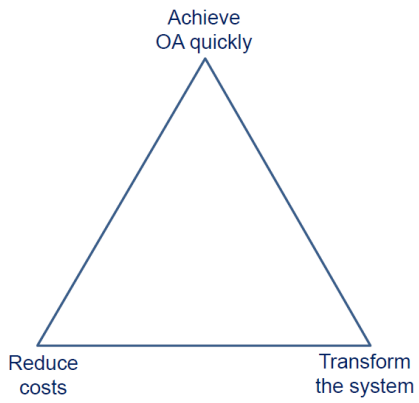
La oss kaste et blikk på utviklingen av transformative avtaler internasjonalt, før vi ser nærmere på effekten av de nye konsortieavtalene ved UiT.

Det vitenskapelige publiseringsmarkedet er dominert av fire akademiske forlag; Elsevier, Wiley, Springer Nature og Taylor & Francis. Innenfor samfunnsvitenskapene utgir disse fire forlagene 70 % av alle artikler som publiseres årlig og innenfor naturvitenskapene utgir de 50 % av årlige artikler. Humaniora er kommersielt

mindre interessant og de fire største forlagene har mindre innflytelse innenfor dette fagfeltet (Røstvik, 2017). Fra bibliotekenes ståsted har det blitt vanlig praksis å abonnere på store abonnementspakker hos de store og mellomstore forlagene, hvor prisen har vært basert på institusjonenes historiske abonnement. De store forlagene har dermed vunnet stadig større markedsandeler med lukkede kontrakter og kontinuerlig høy prisstigning (Bergstrom, Courant, Mcafee & Williams, 2014).

Max Planck Digital Library publiserte i 2015 en rapport som slo fast at det er nok midler i det eksisterende abonnementssystemet til å gå over til en publiseringsbasert økonomi (pay-as-you-publish), men at det kun er bærekraftig dersom abonnementsutgiftene konverteres til å dekke publisering. Modellene må fokusere på åpne og bærekraftige publiseringskostnader og ikke være en omregning av eksisterende abonnementsutgifter. Forlagsavtaler som ser abonnement og publisering i sammenheng er viktige verktøy for å åpne opp abonnementsmidlene, men de er kun et steg på veien (Schimmer, Geschuhn & Vogler, 2015). I høringsrapporten til de norske nasjonale retningslinjene trekkes transformative avtaler fram som et sentralt virkemiddel i overgangen til åpen publisering, men det presiseres også at mange nok må inngå slike avtaler for at det skal få effekt. Samtidig advares det mot faren for å havne i en ny monopolsituasjon (Brekke, 2016). Tidlige erfaringer fra Storbritannia og Sverige har vist at OA-andelen øker ved transformative avtaler, men det gjør også kostnadene (Kungliga Biblioteket, 2017; Pinfield, Salter & Bath, 2017). Lawson (2017) frykter at de store forlagene vil fortsette å beholde sine markedsandeler ved å utvikle innovative avtaler i framtiden.

I arbeidet for åpen tilgang til forskning må ofte ulike mål og prioriteringer balanseres opp mot hverandre. Prioriteringene vil ha effekt på utformingen av politiske strategier og virkemidler. Ifølge Pinfield (2019) er det alltid tre ulike mål som gjør seg gjeldende; oppnå åpen tilgang raskest mulig, endre systemet eller redusere kostnadene. Mange aktører forsøker å nå alle disse målene, men vekt-legger de ulikt, noe som kan illustreres gjennom triangelet i Figur 2.



Figur 3: Stephen Pinfield illustrerer prioritering av ulike mål innen arbeidet for åpen tilgang til forskning.

Avtaler med de tradisjonelle forlagene tillater forskerne å fortsatt publisere i de samme kjente kanalene samtidig som de oppfyller forskningsfinansiørers krav om åpen tilgang. Björk (2016) mener at den raskeste veien til åpen tilgang vil være ved at de store tradisjonelle forlagene konverterer sin portefølje, og at avtaler mellom bibliotek og forlag vil spille en stor rolle. Han uttrykker samtidig skepsis til at bibliotek vil klare å redusere forlagenes maktposisjon og mener at målet om åpen tilgang er overordnet fokuset på å reparere et dysfunksjonelt marked.

Og så kom Plan S

I 2018 kom initiativet Plan S. En rekke forskningsfinansiører, cOAlition S, stiller krav om at all forskning de har finansiert fra og med 2021 skal publiseres åpent tilgjengelig i åpne tidsskrift, plattformer eller arkiv. Publisering i hybridtidsskrift tillates kun dersom tidsskriftet er en del av en transformativ ordning. Kravet til transformative ordninger er at de er midlertidige overgangsløsninger og de må se abonnements- og publiseringskostnader i sammenheng. I tillegg vises det til retningslinjer for transformative avtaler fra *Efficiency and Standards for Article Charges* (ESAC), som peker på transparens, kostnadsnøytralitet og arbeidsflyter (Geschuhn, [2020]). Plan S har også krav om bruk av Creative Commons-lisensen CC BY (med noen unntak) og transparens i hva som ligger til grunn for publiseringsavgifter (cOAlition S, 2020).

Plan S-initiativet har ytterligere aktualisert transformative avtaler. Europa ligger i front, og særlig Tyskland har spilt en sentral rolle, først med sin «no deal» med Elsevier og senere modellavtale med Wiley og Springer. Initiativet OA2020 og ESAC har også base i Tyskland, nærmere bestemt ved Max Planck Digital Library. Men også i USA har noen institusjoner inngått transformative avtaler og i Asia har Japan inngått avtale med Cambridge University Press.

Den økende utbredelsen av transformative avtaler får kritikk fra flere av de åpne forlagene. De peker blant annet på at det transformative aspektet er utydelig og lite konkret i dagens avtaler og at de risikerer å forsinke overgangen til full åpen tilgang (Copernicus, JMIR, MDPI, Ubiquity Press & Frontiers, 2020). På den annen side er det flere bibliotekkonsortier som nå også fokuserer på nasjonale avtaler med de åpne forlagene, og ikke bare de tradisjonelle utgiverne. På denne måten jobbes det med en bredere tilnærming for å nå målet om åpen tilgang og samtidig legge til rette for et sunnere marked.

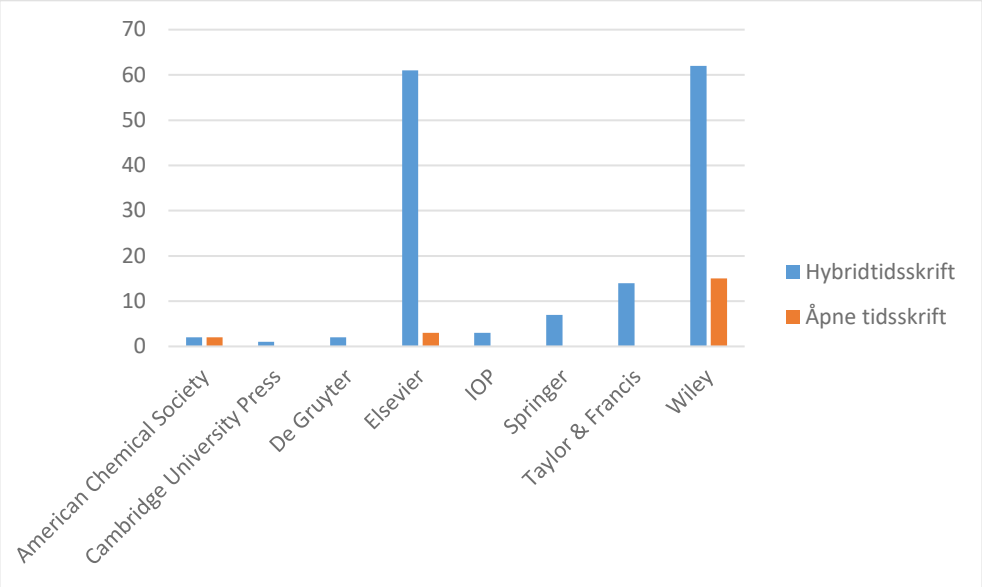
Hva viser UiTs avtaler så langt?

Publisering er et nytt område i avtalene mellom forlag og institusjoner og begge parter har utfordringer med datagrunnlaget. Nødvendig informasjon om korresponderende forfatters institusjonstilhørighet, artikkeltype, inkluderte tidsskrift, med mer, er foreløpig ikke tilgjengelig i våre systemer. Denne utfordringen er ikke unik for Norge, men deles av alle land, konsortier og forlag. Konsortieavtaler

av den typen Unit nå har fremforhandlet på vegne av de norske institusjonene, bidrar dermed til å øke sannsynligheten for at vi i fremtiden vil få bedre datagrunnlag for å studere publiserings-mønstre og innsikt i kostnader ved publisering generelt.

Avtalene er i en tidlig fase, men vi kaster likevel et blikk på den foreløpige statistikken fra UiTs publiseringer i avtalene så langt. For å først gi et bilde av det totale publiseringsvolumet ved UiT, ble det i 2019 publisert 2152 artikler som hadde en medforfatter tilknyttet UiT. Det er ingen enkel måte å stadfeste hvor mange av disse artiklene som hadde korresponderende forfatter fra UiT eller hvor mange artikler som ble publisert i tidsskrift som er knyttet til en av de transformative avtalene. Dermed er det ikke grunnlag for å anslå det fulle potensialet for åpen publisering i avtalene eller hvor mye dette ville utgjort av det totale publiseringsvolumet ved UiT.

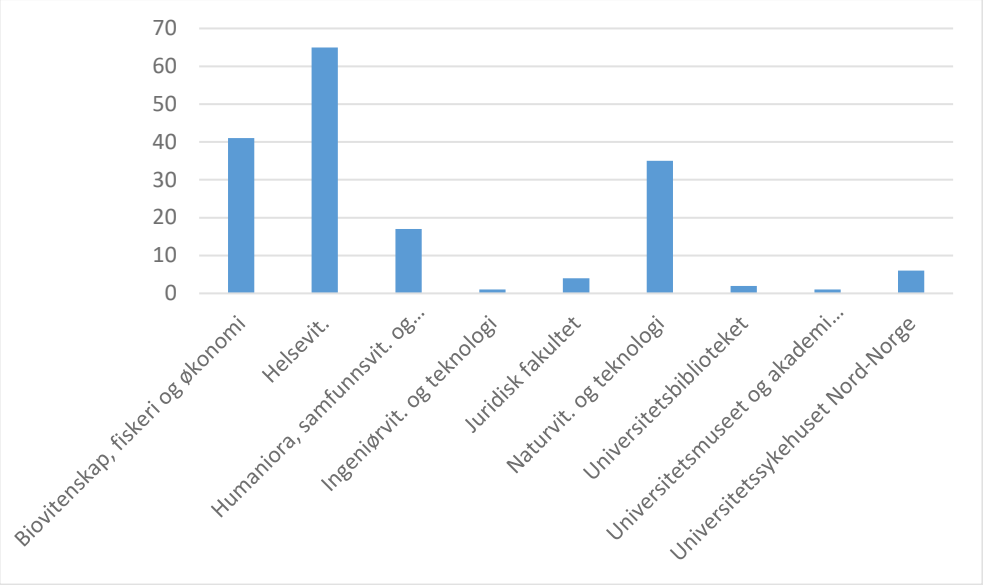
De første avtalene ble implementert i mai 2019. Wiley og Elsevier kom først i gang, mens de fleste andre ble implementert i januar 2020 (se Figur 1 for nærmere informasjon). Siden oppstart og fram til og med mars 2020, har totalt 172 artikler blitt publisert åpent under avtalene.



Figur 4: Antall artikler publisert åpent under avtalene fra mai 2019 til og med mars 2020, fordelt på hybridtidsskrift og åpne tidsskrift.

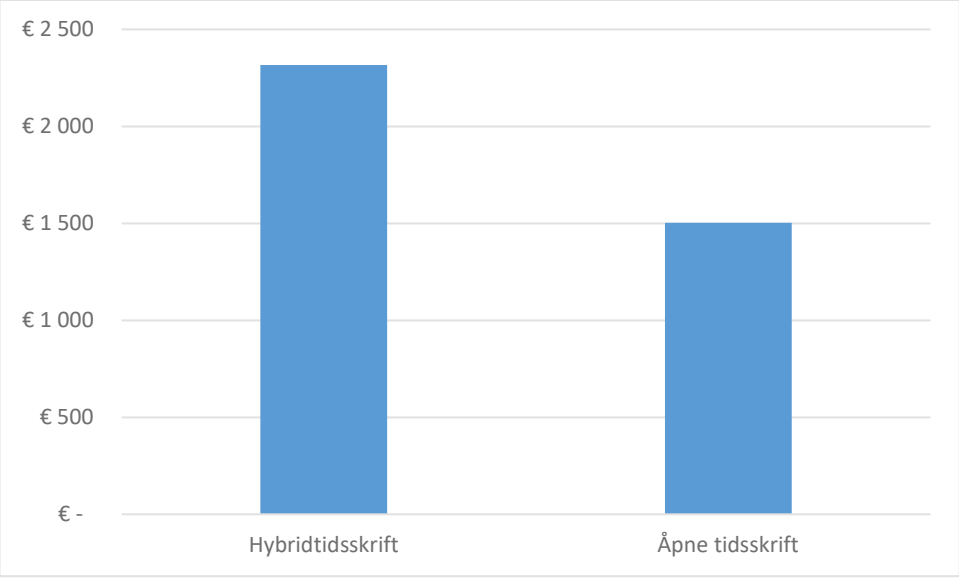
Flest artikler er publisert i hybridtidsskrift. Det er kun Wiley, Elsevier, Cambridge University Press og American Chemical Society som også inkluderer sine åpne tidsskrift i avtalene.

Publisering per fakultet viser at fakultet for helsevitenskap har publisert flest artikler, etterfulgt av fakultet for biovitenskap, fiskeri og økonomi. Datagrunnlaget er ennå for lite til å kunne dra noen konklusjoner, men trenden følger tilsynelatende det generelle publiseringsmønsteret mellom fakultetene, med kun noen få variasjoner. Tidsskriftporteføljen i avtalene vil ha innvirkning på publiseringsstatistikken.



Figur 5: Antall publiseringer som er blitt åpent tilgjengelig i kraft av konsortievtalene, brutt ned på fakultet.

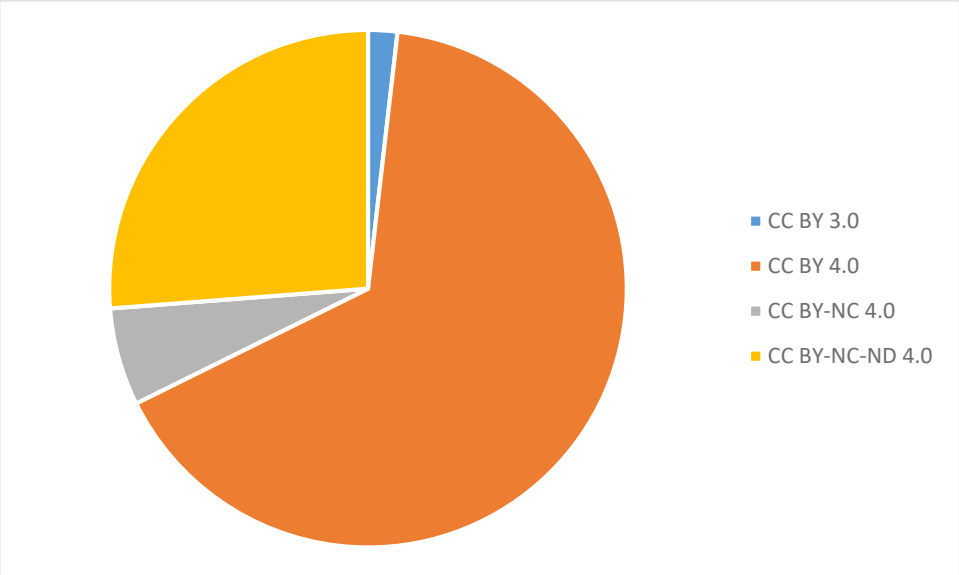
Article Processing Charges (APC) utgjør en mer eller mindre reell størrelse i avtalene. Noen avtaler baserer seg på forlagets reelle APC-priser, mens andre har en kalkulert eller standard APC-pris tilgjengelig. Ut fra dette kan vi anslå snitt APC-pris på UiTs publiserte artikler, og denne bekrefter at APC er høyere i hybridtidsskrift enn i de åpne tidsskriftene som er inkludert i avtalene.



Figur 6: Snittpris på APC inkludert rabatter i hybridtidsskrift og åpne tidsskrift.

De mest vanlige Creative Commons-lisensene i avtalene er CC BY og CC BY-NC-ND eller CC BY-NC, men det kan variere fra tidsskrift til tidsskrift innenfor avtalen. CC BY er lisensvarianten som sikrer best mulighet for spredning og gjenbruk og UiT anbefaler forskerne å velge denne lisensen så langt det er mulig. Det er også lisensvarianten som frontes av Plan S. I de tilfeller hvor forlagene tilbyr en lisensvariant

som er ikke-kommersiell (NC), vil det være forlaget som beholder de kommersielle rettighetene og ikke forfatteren selv.



Figur 7: Valg av Creative Commons-lisenser, basert på 164 artikler av 186 totalt. Lisens for de resterende artiklene er foreløpig ikke tilgjengelig.

Så langt, alt vel?

I tråd med de nasjonale retningslinjene har altså Unit forhandlet frem åpen publisering som en del av avtalene med de tradisjonelle forlagene. Siden 2017 har hele elleve avtaler kommet på plass. Det er for tidlig og for ufullstendige data til å trekke noen konklusjoner ut fra publiseringstallene, men så langt peker det i retning av at flere artikler nå publiseres åpent, CC BY er den mest foretrukne lisenstypen og publiseringsmønsteret følger fagmiljøenes generelle mønstre. Tallene peker også på at avtalene bidrar til økt åpen publisering i mer kostbare kanaler.

Regjeringen stiller krav om transparens og kostnadsnøytralitet, og Plan S har også en rekke krav til både transformative ordninger og åpen publisering for øvrig. For de norske nasjonale avtalene er avtalevilkårene åpne og lisensene er lagt ut på Units nettsted openaccess.no. Noen av avtalene kan sies å være kostnadsnøytrale, det vil si at prisstigning ikke overskrider det som tidligere har vært normalt. Men for flere av avtalene har prisstigningen vært høyere enn normalt, hovedsakelig i avtaleperiodens første år. Foreløpig bidrar dette til å forsterke de store forlagenes markedsposisjon og spiser ytterligere av bibliotekenes trange budsjetter. Muligens er det akseptabelt, all den tid dette skal være midlertidige overgangs-løsninger. Avtalene legger opp til å dekke all publisering av norske korresponderende forfattere og legger til rette for overgang til en publiseringsbasert økonomi. På den annen side, og kanskje helt naturlig, er det lite konkret hva som vil skje etter endt avtaleperiode. Spørsmål som melder seg, er hvordan bibliotekene skal håndtere

materiale som fortsatt ligger bak betalingsmur? Så langt er ikke hele tidsskriftporteføljen til forlagene inkludert i avtalene og med mindre det blir global oppslutning bak slike avtaler om åpen tilgang, vil det trolig fortsatt være behov for løsninger som sikrer lesetilgang. Og hva med de mindre utgiverne? Det er minst like krevende å forhandle avtaler med små som med store forlag.

Avtalene muliggjør åpen publisering i kanaler som er godt representert i det norske registeret over vitenskapelige publiseringskanaler, og nettopp her ligger de sterkeste insentivene for hvor norske forskere velger å publisere. Avtalene bidrar dermed til økt åpen publisering, men kan de bidra til å endre systemet eller redusere kostnadene? Kanskje det nå er viktigere enn noensinne å ha flere tanker i hodet samtidig. I 2017 foreslo det svenske Kungliga Biblioteket at kommersielle aktører burde erstattes av publiseringsløsninger eid av akademia. I Sør-Amerika er dette allerede den utbredte modellen og i 2018 ble initiativet AmeliCA lansert. Under slagordet «Åpen kunnskap uten kommersielle formål eid av akademia» bygger AmeliCA infrastruktur for åpen publisering som på rekordtid er tatt i bruk av portugisisk- og spansk-talende land over hele Sør- og Mellom-Amerika. Kanskje vi burde se oftere mot sør?

Referanser

Bergstrom, T. C., Courant, P. N., Mcafee, R. P. & Williams, M. A. (2014). Evaluating big deal journal bundles. *Proceedings of the National Academy of Sciences, 111*(26), 9425. https://doi.org/10.1073/pnas.1403006111

Björk, B.-C. (2016). The open access movement at a crossroad: Are the big publishers and academic social media taking over? *Learned Publishing, 29*(2), 131-134. https://doi.org/10.1002/leap.1021

Brekke, T. (2016). *Nasjonale retningslinjer for åpen tilgang til forskningsresultater: Rapport til Kunnskapsdepartementet 14.06.2016*. Hentet fra https://www.regjeringen.no/contentassets/72e9794a183647e5b53ec39ba8cf516a/rapport-nasjonale-retningslinjer-for-apen-tilgang-til-forskningsresultater.pdf

cOAlition S. (2020). Plan S Making full and immediate Open Access a reality. Hentet 29.03 2020 fra https://www.coalition-s.org/

Copernicus, JMIR, MDPI, Ubiquity Press & Frontiers. (2020, 31.3.). Current Transformative Agreements Are Not Transformative: Position Paper – For Full, Immediate and Transparent Open Access. Hentet fra https://frontiersinblog.files.wordpress.com/2020/03/position-statement-transformative-agreements.pdf

Geschuhn, K. ([2020]). ESAC efficiency and standards for article charges - Guidelines for Transformative Agreements. Hentet 29.03 2020 fra https://esac-initiative.org/about/transformative-agreements/guidelines-for-transformative-agreements/

Kungliga Biblioteket. (2017). *Utvärdering av offset-avtal – delrapport 2: Springer Compact och Institute of Physics*. Kungliga biblioteket. Hentet fra http://openaccess.blogg.kb.se/files/2017/09/Utva%CC%88rdering-av-offset-avtal-SC-och-IOP-delrapport-2.pdf

Kunnskapsdepartementet. (2017). Nasjonale mål og retningslinjer for åpen tilgang til vitenskapelige artikler. Hentet 12. oktober 2017 fra https://www.regjeringen.no/no/dokumenter/nasjonale-mal-og-retningslinjer-for-apen-tilgang-til-vitenskapelige-artikler/id2567591/

Lawson, S. (2017). *Report on offset agreements: Evaluating current Jisc Collections deals. Year 2 – evaluating 2016 deals*. Hentet fra https://figshare.com/articles/Report_on_offset_agreements_evaluating_current_Jisc_Collections_deals_Year_2_evaluating_2016_deals/5383861

Pinfield, S. (2019, 22.10.2019). *The role of research libraries in supporting open science: The situation in UK and beyond*. Innlegg presentert ved Høstseminar Norsk Fagbibliotekforening Trondheim, Trondheim. Abstract hentet fra https://nff-trondheim.norskbibliotekforening.no/2019/09/12/hostseminar-22-oktober-2019/

Pinfield, S., Salter, J. & Bath, P. A. (2017). A "Gold-centric" implementation of open access: Hybrid journals, the "Total cost of publication," and policy development in the UK and beyond. *Journal Of The Association For Information Science And Technology, 68*(9), 2248-2263. https://doi.org/10.1002/asi.23742

Røstvik, C. M. (2017). Publisering, penger og prestisje: Historien om hvordan akademisk publisering ble kommersialisert. *Nytt Norsk Tidsskrift, 34*(03), 301-309. https://doi.org/10.18261/issn.1504-3053-2017-03-07

Schimmer, R., Geschuhn, K. K. & Vogler, A. (2015). Disrupting the subscription journals' business model for the necessary large-scale transformation to open access: A Max Planck Digital Library Open Access Policy White Paper. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AJRG23.v1

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Support for Good Peer Review in OJS-based Journals: A Library Publisher's Perspective[1]

Aysa Ekanger and Solveig Enoksen

Abstract

How can a library publishing service with limited resources help editorial teams of peer-reviewed journals in their work? This paper focuses on the technical aspects of the peer review workflow that, if set up and adhered to properly, can contribute to improving the standard of the peer review process – and to some degree also the quality of peer review. The discussion is based on the work done at Septentrio Academic Publishing, the institutional service provider for open access publishing at UiT The Arctic University of Norway.

Introduction

Septentrio Academic Publishing is a service offered by the University Library of Tromsø to open access journals and series that are associated (through editors or scope) with UiT The Arctic University of Norway.[2] The service is part of the library's and the university's commitment to the open access mission: Septentrio offers its support for free and strives for its journals to have good quality and be visible in the open access infrastructure. One aim is that all peer-reviewed journals in Septentrio are indexed in the Directory of Open Access Journals, DOAJ. As of May 2020, Septentrio publishes ten peer-reviewed journals in a variety of disciplines (eight of them in the DOAJ) and eight series that are not formally peer-reviewed.

A handful of library employees are involved in running the publishing platform (Open Journal Systems from the Public Knowledge Project), providing technical support to the users, and giving advice to the

[1] This paper is a result of our presentation about support for peer review given at the midterm meeting of the IFLA Library Publishing Special Interest Group that took place at OsloMet in March 2020. IFLA stands for The International Federation of Library Associations and Institutions, https://www.ifla.org/.

[2] As this paper is published in a Ravnetrykk issue in honour of Stein Høydalsvik, it is appropriate to mention that Stein was one of the founders of the publishing service – and establishment plans for the service were mentioned in Høydalsvik (2002), https://doi.org/10.7557/15.3844.

https://doi.org/10.7557/15.5499

editorial teams about best practice in various aspects of the publishing process, such as peer review. The amount of (wo)man-hours that we can invest in Septentrio work is limited – as all of us have other tasks. How can a library publishing service with limited resources help the editorial teams in their work? Our goal is to make the publishing process smoother, through appropriate workflows set up according to a journal's needs on the journal platform and through relevant information provided to editors, reviewers and authors at crucial points.

This paper focuses on the technical aspects of the peer review workflow that, if set up and adhered to properly, can contribute to improving the standard of the peer review process – and to some degree also the quality of peer review.

Support for peer review

The publishing software Open Journal Systems (OJS) offers a good workflow for the peer review stage of the publishing process, with a number of options that allow to model the peer review process in accordance with what a specific journal needs. However, not all our journals use OJS for peer review: some editors prefer the more familiar method of managing peer reviews via email, outside of OJS.

When we try to explain to editors why they should use OJS for peer review, we usually point to the benefits and importance of keeping manuscript history in one place.[3] Improvement of the standards of the peer review process is another reason for why peer reviews should be handled inside OJS – instead of via email.

So, how can we as technical support staff help editorial teams with the peer review process? Peer review-related issues that can be affected by technical configurations, can be roughly divided into two groups: (adherence to) ethical standards and technicalities pertaining to blind review (single-blind or double-blind).

Authors, reviewers and editors should be aware of these standards and technicalities, it must be technically possible for them to comply with the requirements – and they probably also need to be reminded of them at various stages in the publishing process. Editors also have the added responsibility of making sure that authors and reviewers are following all of these standards and technicalities.

The Public Knowledge Project provides openly available user guides for the publishing workflow in OJS, where almost every step of the

[3] During the publication process, the platform provides information on when review reports are due and sends automatic reminders to users when an action needs to be taken – so that it becomes easier for an editor to manage a journal. In addition, overview over both active and published or archived submissions is important when there are multiple editors or section editors involved, when editorial teams change, or when guest editors come in to publish a specific issue.

editorial process or the review process is described in detail. An important task of the support staff at Septentrio is to make editorial teams and other users aware of the OJS functionality, help to set up the review process in a way that excludes slip-ups, and make adherence to various guidelines and requirements easier by making them visible at various steps in the publishing process. The Septentrio team have also put together a number of guides on the publishing process, openly available to editors, reviewers and authors on the Septentrio website.

Ethical standards

There are guidelines for ethical standards in peer review from various associations. A good place to start for both the library support staff and journal editors are Ethical Guidelines for Peer Reviewers from the Committee on Publication Ethics (COPE)[4], as well as the Principles of Transparency and Best Practice in Scholarly Publishing[5]. Journal editors also need to be aware of discipline-specific guidelines relevant to their specific journal.

COPE advises reviewers to follow the journal's instructions for peer review – this emphasizes how important it is for a journal to have clear guidelines available for reviewers.

In November 2019, in connection with the 14th Munin Conference on Scholarly Publishing organized by the university library, COPE was invited to hold a workshop on the standards of peer review. Editors of Septentrio journals were personally invited to participate in the workshop, without admission fees. In 2020, Septentrio is planning to help its journals to assess whether their peer review processes are carried out in accordance with the standards of COPE.

The ethical standard issues that can be taken care of with the help of technical adjustments, include:

- ensuring against reviewer bias and competing interests
- accommodating for appropriate reviewer feedback and dealing with inappropriate reviews
- dealing with suspected ethics violations on the part of the authors

Review feedback should be unbiased – not influenced by the reviewer's possible competing interests, such as a close connection to the author or, on the contrary, being part of a competing research group.

---

[4] COPE Council (2017). Ethical guidelines for peer reviewers. https://doi.org/10.24318/cope.2019.1.9

[5] COPE, DOAJ, OASPA, WAME (2018). Principles of Transparency and Best Practice in Scholarly Publishing. Version 3. https://doi.org/10.24318/cope.2019.1.12

A journal should have a clear conflict of interest disclosure policy displayed in the reviewer guidelines on the journal's website. In addition, in the set-up of the review workflow in OJS, a journal should choose to make the conflict of interest disclosure policy visible to reviewers when they log onto the publishing platform. The workflow settings also allow a journal to choose to request that reviewers submit a competing interests statement.

Reviewers are supposed to give appropriate feedback. This, according to COPE, consists of a fair, honest, and unbiased assessment of the strengths and weaknesses of the manuscript. The journal will of course be reminding reviewers about this in its guidelines, but there is also a tool in OJS that journals can use – namely review forms.

A review form consists of predefined questions or statements that reviewers can react to in a variety of ways: by filling in a text box, by checking off several checkboxes, by choosing one radio button or an item from a dropdown menu. It is possible to set up different review forms for different types of papers that a given journal accepts (articles, essays, squibs). By setting up review forms, a journal ensures that reviewers address the issues they need to address. Another benefit of review forms is that they help to standardize the review process and make it easier for editors to compare feedback on the same manuscript from different reviewers.

However thoroughly the review workflow may be set up, there is always a risk of inappropriate reviews. You may have heard about cases where e.g. the author's intellect is put into question, or there is gender bias. It may also happen that the reviewer clearly does not know the subject, so the review feedback is practically worthless. There are guidelines on how to deal with inappropriate reviews (from COPE, or from disciplinary associations), but editors should also know about technical options OJS provides them with in such situations.

If an inappropriate review has been submitted in OJS, an editor can *unconsider* it – the author will then not receive this review feedback. For future reference, the editor can then rate the reviewer and include notes with an explanation of the rating – the rating and notes will not be visible to the reviewer.[6]

While reviewing manuscripts, reviewers not only have to adhere to ethical guidelines themselves, but also to make sure that authors do so as well. What if a reviewer suspects a violation of ethics in the research process presented in the manuscript or detects a case of plagiarism?[7] In OJS, a reviewer can inform the editors of a suspected

[6] The option of reviewer rating is always there – not only in the case of inappropriate reviews. The rating system is designed to make it easier for editors to choose reviewers from the database on the publishing platform.

[7] As a member of Crossref, Septentrio has access to the plagiarism-checking service Similarity Check, powered by iThenticate. All journals at Septentrio

violation without including the author(s) in the communication. There are different options for doing this. A reviewer can express her concerns in the "For editor only" textbox when submitting the review. A reviewer can also use the "Add discussion" option in the system (choosing the editor as the only other participant in the discussion) during the review process, before submitting the final review feedback. Using OJS for this type of communication also ensures that suspicions of ethics violations are saved for future reference, rather than being lost in an editor's email inbox.

Blind review technicalities

Of our ten peer-reviewed journals, two practice single-blind review, and eight – double-blind review. The fact that the identities of the reviewers and the authors cannot be disclosed to the other party means more things for authors, reviewers and editors to remember.

An author may inadvertently reveal her identity in the manuscript, replication data, the Conflict of Interest statement, and by making a copy of the manuscript available on a preprint server before submission. A reviewer may slip up and identify herself in the review report. Both author and reviewer may leave identifying traces in file properties and comments. OJS has built-in functionality for ensuring anonymity of reviewers and authors: when the blind or double-blind review option is chosen in the set-up of the review workflow, the system controls that authors and reviewers do not get access to identifying metadata. In addition, authors and reviewers are presented with the link to "Ensuring blind review" guidelines (pre-set by the OJS developers) at the relevant stages of the publishing process. Journals should also provide information about how to ensure anonymity in their Author Guidelines and Reviewer Guidelines, and set up the submission and review process in a way that precludes unwanted identification.

The author needs to make sure that there is no identifying information in the manuscript. This includes names and affiliations, acknowledgements, and the phrasing of self-citations. If a journal provides submitting authors with article templates, it may be preferable not to include the field for author name – so that there is one less thing for an editor to remember to take care of before sending a manuscript off to a reviewer. Authors – and also reviewers, if they are uploading any documents as part of their review reports – must remember to remove identifying information from file properties (the "last edited by" and "author" fields) and from any comments that have

---

can use the service (and the ones indexed in the DOAJ have to use it), and the library covers the Similarity Check fee for the journals. It is the editors' responsibility to run the submitted manuscripts through plagiarism check before reviewers are assigned.

been made using the “insert comment” functionality in their text-processing program.

If replication data are submitted to a data repository and made available to the journal’s reviewers, the author needs to inform the data repository curators that the dataset is meant to be made available to reviewers in a journal that practices double-blind peer review. The curators will then take the necessary precautions, e.g. by providing reviewers with a copy of the dataset that does not contain identifying information in its metadata or ReadMe-file.

When submitting to a journal using double-blind reviews, authors need to be careful not to include the Conflict of Interest (CoI) statement in the manuscript. It is advised that the full statement is sent as a Comment to the Editor during the submission in OJS, i.e. outside of the manuscript. The author should also provide a short version of the CoI (e.g. only stating *whether* there is a conflict of interest) in the Comment to the Editor (which can be forwarded to the reviewer by the editor) or included in the submitted manuscript (to be substituted by the full CoI before publication).[8]

Ensuring double-blind review may be challenging if the author's manuscript is openly available as a preprint. A journal's editorial team need to think in advance about their position on this issue: do they want to accept manuscripts for which preprints are available somewhere? The answer to this question should be “yes”: preprints contribute to the transparency and efficiency of scholarly communication, and there are fewer and fewer journals who have restrictive policies with regard to preprints. There is then not much for a journal to do except inform its authors and reviewers (in Author Guidelines and Reviewer Guidelines) that preprint availability is not a tragedy – even if the journal standardly operates with double-blind review. The author has to accept that double-blind procedure will not be guaranteed in this case – and reviewers should not reject an invitation to review a manuscript merely due to the availability of a preprint.

A reviewer may accidentally reveal her identity in the review report, e.g. by signing her name or by referring to something that can identify her. In OJS, the default, free form for a review report is separated into a field that is addressed to both editor and author, and another one that is addressed to the editor only. When forwarding a review report to an author, an editor should ensure that no identifying information is included in the forwarded part. A pre-defined review form – that we mentioned above in connection with appropriate review feedback – reduces the risk of identifying information being included in a review report. A review form does not contain the fields addressed to the

[8] A reviewer’s CoI statement will not be visible to the author, as long as the review process is marked as blind or double-blind in OJS.

author and/or the editor – so there is less possibility that a reviewer may confuse the fields. Also, a review form – as it consists of a set of predefined questions and statements – may help to keep reviewers focused.

To sum up

There are standards and technicalities in the peer review process that authors, reviewers and editors need to be aware of, be able to comply with and be reminded of. The editor has a central role in monitoring adherence to ethical standards and anonymization technicalities.

A library publishing service can help the editors make the review process as smooth as possible, with a focus on the importance of adhering to ethical standards and ensuring the anonymity of the review process. The library should offer training in the use of OJS to the editors and may consider, as in our case, compiling a set of editorial resources. Producing editorial resources specific to your publishing service may seem unnecessary when there are good user guides made available by OJS and COPE. However, workflows differ – on publisher and journal level – so it may be useful for editors to have access to step-by-step guides that are compiled specifically for their needs.

When the technical part – including easily accessible guidelines – is set up properly, there are fewer possibilities for authors and reviewers to make mistakes and the review process is standardised for all submissions. This, in turn, improves the overall quality of the review process. A good quality of the peer review process from start to finish is something all journals should strive for, and, as shown here, a library publishing service can play an important role in achieving this.

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

The costly prestige ranking of scholarly journals

Bård Smedsrød and Leif Longva

Abstract

The prestige ranking of scholarly journals is costly to science and to society. Researchers' payoff in terms of career progress is determined largely from where they publish their findings, and less from the content of their scholarly work. This fact creates perverted incentives for the researchers. Valuable research time is spent in trying to satisfy reviewers and editors, rather than spending their time in the most productive direction. This in turn leads to unnecessary long time from research findings are made until they become public. This costly system is upheld by the scholarly community itself. Scholars supply the journals with time, serving as reviewers and editors without any paycheck asked, even though the bulk of scientific journals are published by big commercial enterprises enjoying super profit margins. The super profit results from expensive licensing deals with the scholarly institutions. The free labour offered, on top of the payment for the licensing deals, should be viewed as part of the payment to these publishers – a payment in kind. Why not use this as a negotiating chip towards the publishers? If a publisher asks more than acceptable for a licensing deal, rather than walk away with no deal, the scholarly institutions could pull out all the free labour offered by reviewers and editors.

Major loss of efficiency and productivity

It is commonly thought that scholarly journals with high rejection rates have the highest prestige among the journals. A tough filtering regime is held by many as the foundation to maintain a high-quality level of the journal: only the best of the best manuscripts will end up being published in the most prestigious journals. Based on this reasoning a hierarchy of prestige ranking among the journals is formed (1), a hierarchy that by many is viewed to be of great value (although disputed (2)), helping the scientists select what to read and easing the ranking of job and grant applicants. But what are the costs induced by this model of prestige ranking?

A manuscript submitted to a prestigious journal, unless immediately rejected, is normally subjected to a lengthy reviewing process which may take several months, or even years before the manuscript is published and accessible. The authors must use valuable time to

https://doi.org/10.7557/15.5507

polish the manuscript and are frequently told by the reviewers to carry out additional analyses that often do not add much, if at all, to the reliability and quality of the results (3). The main reason why this is costly and counter-productive is that most of the prestige journals are not run as open access journals. The common and costly model is that the results presented in the submitted manuscripts are normally kept totally hidden during the lengthy period from submission to publishing, and next only accessible behind expensive access bars after publication.

The long period of inaccessibility of the article, along with the polishing and additional experiments that may at best result in marginal improvements, represents a major loss of efficiency and productivity and thus represents an economic cost to research and society. It is important also to keep in mind all the manuscripts that are submitted to journals but rejected. These may also be kept in the editorial and review process for quite a lengthy time, before rejection. And a new round of submission and editorial and review process starts.

Why is it that researchers keep sending their manuscripts to journals operating like this, and hiding their findings away from fellow researchers and the society for a lengthy time period? The answer is obvious: Researchers get their pay-off from publishing in prestigious journals. Hence, while the decreased research efficiency caused by the prestigious journals is a cost to society it is not perceived by the individual researcher as a cost. If pay-off was not tied to the prestige of the publishing channel, authors could choose publishing channels that are more efficient in disseminating their new research findings, which would remove the entire problem of keeping research results needlessly hidden during a lengthy review process.

Alternative models

Alternative publishing models have been described and also launched. Examples: The article is published openly, after a control ensuring that the research is based on well-founded methods. Or the article may be available in a preprint server. The review process may thereafter continue openly, both by reviewers selected by the journal, and by other interested peers. In this way the pieces of new knowledge may come to use without the delay caused by journals (4, 5, 6, 7).

It is interesting to note how the recent COVID-19 outbreak has spurred publishing via preprint servers and other ways of sharing research results (8). The rapidly spreading pandemic is a serious threat to public health, society infrastructure and world economy, and research results that may give clues about how to combat the pandemic, developing therapeutic remedies, and reduce its adverse effects on our society are desperately needed ASAP. In this situation, it is obvious that publishing in prestige journals with a lengthy review process is not only counterproductive and inefficient use of public

money. It is also unethical. It may be argued that rapid publishing of research results prior to peer reviewing may increase the risk that serious errors in methods or conclusions in pre-reviewed manuscripts may lead to serious adverse results – in particular if the results are to be used in clinical settings. To avoid this, it is of course utterly important that readers of pre-reviewed publications understand that publishing on preprint servers is to be followed up by a review process. Researchers who read these manuscripts must be aware that they themselves will have to read the manuscript critically and thus contribute to doing peer reviewing. And after all, this is what peer reviewing is all about: competent and interested colleagues critically scrutinizing each other's manuscripts. This may be the way of publishing in the future: it will ensure maximal speed of results dissemination, not only in times of desperate need for results to combat pandemics. In fact, all research deserves to be published ASAP, independently of how urgently the results are needed.

Another advantage of speeding up the publication process by shifting to a pre-review publishing regime is to reduce the possibility that researchers are being scooped during the often lengthy review period that the manuscript is hidden from the general public. It is no secret that many researchers fear the possibility that their findings and ideas may be used by reviewers, who are often their competitors. By first posting the manuscript on a publicly accessible preprint server, for an open peer review to take place, the publishing researcher can feel safe that he or she will be registered as the first to publish results and/or ideas that might otherwise be unrightfully taken by others during the lengthy review process that most journals make use of today.

A shift towards a publishing model along these lines would result in a faster dissemination of research findings, avoiding the costly system of hiding research in a lengthy review and editing process. And a prestige ranking of the journals would cease, if rejection rates no longer would bear any meaning. In addition, this publishing model represents a major driver towards open science.

However, successful researchers who have won their reputation by making every effort to publish in high ranked journals will dislike this change in the publishing model, and they may thus constitute a barrier to change. But can we afford to maintain the existing model, which is both counterproductive to research and very costly to society?

But the journal hierarchy is a help for the reader, isn't it?

What about the readers of scientific articles – don't they benefit from the journal hierarchy? After all, with the enormous rate of publishing there is no way one will have the time to read but a fraction of what is published anyway. So, help and guidance in selecting literature worth reading is something any researcher would highly appreciate.

Hence it may appear logical that reading “the best journals” would be tantamount to reading the best articles and thus avoiding wasting time looking through journals that are deemed less prestigious.

This conclusion may not appear so obvious when considering how the prestigious journals select the articles to publish. Articles with the most sensational findings are preferred, since these help maintaining a high journal impact factor. These articles are not necessarily the ones most useful to the research of peers and colleagues. Moreover, internet-based communication has come a long way, offering far more precise (and speedy) signaling methods than the journal title to find literature of high relevance and value to the individual researcher. The potential for new selection methods is great, if only we can avoid the access barriers that still dominate in journal publishing.

Evaluation of the researcher

A generally accepted presumed advantage of maintaining the journal hierarchy has until now been to contribute to the assessment of the excellence of researchers, in applications for research grants and job promotions. Researchers are thus evaluated not based on what they have accomplished in their research, but rather on where they have published. The problem with this – as well as an effort to remedy the problem – is clearly described in the DORA declaration (9). Among several measures to reform research assessment, the declaration highlights that the scientific content of a paper is much more important than traditional publication metrics or the name of the journal in which it was published. This evaluation system is the main reason why the researchers find it beneficial to work intensively to have their manuscripts published in high prestige journals and compete with each other instead of cooperating. This competition may drive the researchers to work hard. But it also adds to the loss of efficiency and productivity by keeping the findings hidden for a lengthy period and forcing researchers to keep kneading their papers instead of moving on with their research in the most fruitful direction. So alternative methods of researcher assessment are needed (10).

The peer review task

It is frequently maintained that the prestigious journals, in addition to attracting good manuscripts, also attract the best reviewers and editors. Based on this it is argued that these good contributions to research will be lost if the journal hierarchy is abolished. But wait! These reviewers are researchers who commonly are employed by universities and research institutions and in most instances work for free for the journals. Even if the entire journal hierarchy were to disappear the good reviewers and editors would still be around. Their skills and valuable work effort would still contribute importantly to research, for the good of society. Clearly, it is the researchers, not the journals who assess the research quality. And it is the employers of the researchers who grant them the possibility to use their work

hours and skills to serve as reviewers and editors. It is remarkable that these employers pay no attention whatsoever to how and for which journals their researchers perform the pivotal quality control of their peers' manuscripts.

In need of change

The strong incentives for the individual researcher to dance to the tune of the prestigious journals, call for measures so academia may regain control. A drastic action that is presently discussed and even put into effect in some countries is to discontinue the subscription agreements with the big international publishing houses that refuse to lower their price charges and/or to move towards an Open Access publishing model (e.g. as laid out in Plan S). Just as important, it is high time that the university sector takes control over the review activities performed by their employed researchers. One way to attain this goal may be for the university sector to issue lists of journals for which review tasks should not be done. It is the peer review work that gives the journals their quality stamp. By controlling this the universities will take back the control of the publishing, enabling new and more healthy goals to be specified for the publishing policy of the future. Major goals are to attain open research and publishing processes and reach agreements with the publishing houses to pay only for the services they actually provide. The course of action needed to do this, i.e. controlling peer review, may mean reduced academic freedom for the researchers, by some limitations on where to do peer review. But using these measures to gain control over the journal hierarchy is necessary to remove the barriers that are economically unacceptable and counteractive to a sound development of research (11).

Moreover, the peer review job, and also the job as editors, that researchers do for the journals, is normally done without any payment for this highly valued job. This supply of free labour should be viewed as part of the payment from the academic institutions to the publishers. It is a payment in kind. So, if the academic institutions are less than pleased with a publisher's offer in terms of price and services (including willingness to move towards Open Access publishing), the institutions may withdraw this extra in kind payment they ship to the publisher in question. But in order to do so, the institutions need the power to instruct their employees on where to do peer review. This could be a very powerful threat. Where would the prestige of a publisher's journals go, if access to reviewers and editors became limited?

The present system of scholarly publishing, designed for Gutenberg's 500 years old technology, is more than ready for a major revision. The big international publishers are not the prime movers here. They have no wish to change the system that so far has given them enormous profit. Instead, the universities ought to have the motivation and possibility to develop more healthy models for scholarly publishing. And the means to do so are available as shown

recently by University of California faculty (12). There, a prominent group of faculty members announced to step away from the editorial boards of scientific journals published by Elsevier until the publishing giant agrees to restart negotiations and agree to a fairer deal with respect to access and charges.

So, our claim is that academia does have the means to improve the publishing model. The benefits include reduced costs for academia and the society at large, and, not less important, a possibility to move towards a more productive research.

References

1. Brembs, B., Button, K. & Munafò, M. (2013). Deep impact: unintended consequences of journal rank. Front. Hum. Neurosci. 24 June 2013. https://doi.org/10.3389/fnhum.2013.00291

2. Brembs, B. (2018). Prestigious Science Journals Struggle to Reach Even Average Reliability. Front. Hum. Neurosci. 20 February 2018. https://doi.org/10.3389/fnhum.2018.00037

3. Snyder, S.H. (2013). Science interminable: Blame Ben? Proc Natl Acad Sci USA 110(7):2428–2429. https://doi.org/10.1073/pnas.201300924

4. Stern, B.M. & O'Shea, E.K. (2019). A proposal for the future of scientific publishing in the life sciences. PLoS Biol 17(2):e3000116. https://doi.org/10.1371/journal.pbio.3000116).

5. Sever, R., Eisen, M. & Inglis, J. (2019). Plan U: Universal access to scientific and medical research via funder preprint mandate. PLoS Biol 17(6):e3000273. https://doi.org/10.1371/journal.pbio.3000273

6. Tennant, J. (2018). The State of The Art in Peer Review. SocArXiv. May 29. http://dx.doi.org/10.1093/femsle/fny204

7. Green, T. (2019). Is open access affordable? Why current models do not work and why we need internet-era transformation of scholarly communications. Learned Publishing, Volume 32, Issue1, p 13-25. http://doi.org//10.1002/leap.1219

8. Kupferschmidt, K. (2020). 'A completely new culture of doing research.' Coronavirus outbreak changes how scientists communicate. https://www.sciencemag.org/news/2020/02/completely-new-culture-doing-research-coronavirus-outbreak-changes-how-scientists

9. San Francisco Declaration on Research Assessment. https://sfdora.org/read/

10. European Commission, Directorate-General for Research and Innovation (2017): Evaluation of Research Careers fully

acknowledging Open Science Practices. http://doi.org/10.2777/75255

11. Longva, L., Reierth, E., Moksness, L. and Smedsrød, B. (2017). Peer reviewing: a private affair between the individual researcher and the publishing houses, or a responsibility of the university? Journal of Electronic Publishing, Volume 20, Issue 1, 2017. http://dx.doi.org/10.3998/3336451.0020.103
12. Sanders, R (2019): UC faculty to Elsevier: Restart negotiations, or else. Berkeley News, August 7, 2019. https://news.berkeley.edu/2019/08/07/uc-faculty-to-elsevier-restart-negotiations-or-else/

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Seniorprosjektet 2017–2019

Sluttrapport

Lars Figenschou

Sammendrag

God forvaltning og deling av forskningsdata er helt sentralt for å fremme transparens og kvalitet i forskningen. Seniorprosjektet (2017–2019) har tilbudt UiTs forskerseniorer hjelp med tilrettelegging av forskningsdata slik at dataene ikke går tapt. I samarbeid med den enkelte forsker har vi beskrevet datagrunnlaget og tilrettelagt for fremtidig åpen tilgang og publisering, der det har vært aktuelt.

Prosjektbakgrunn og formål

Universitetsstyret vedtok i 2016 digitaliseringsstrategi for UiT og behandlet samtidig en plan for implementering (hhv Sak 33/16 og 65/16). I denne planen lå det konkrete forslag til satsinger ved UiT, og prosjektet «Støtte til UiTs seniorer for tilrettelegging av forskningsdata for digital arkivering og gjenbruk» var en av satsingene som fikk midler.

Prosjektet har hatt som formål å ta vare på forskningsdata fra UiTs forskerseniorer. Dataene som ligger hos seniorforskerne som er på vei ut av akademia, er data som er mest utsatt for å gå tapt både for fremtidig forskning og som dokumentasjon og verifisering av tidligere forskning.

Fremgangsmåte

I perioden fra 2017 til og med 2019 har prosjektet tilbudt UiTs forskerseniorer hjelp med tilrettelegging av forskningsdata for digital arkivering og bruk i videre forskning. Arbeidet har vært organisert i et nært samarbeid med fakultetene når det gjelder hvilke forskere og hvilke forskningsdata som har vært prioritert.

Prosjektet har bistått enkeltforskere fra denne gruppa med tilrettelegging av forskningsdata for trygg og langsiktig lagring med sikte på tilgjengeliggjøring. Avdeling for IT (ITA) har stilt tilstrekkelig lagringskapasitet til disposisjon for forskerne slik at vi har fått flyttet data fra sårbare lagringsmedier til trygg og langsiktig lagring. I tillegg har prosjektet beskrevet datagrunnlaget og tilrettelagt for fremtidig åpen tilgang og publisering, der det har vært aktuelt.

https://doi.org/10.7557/15.5515

Resultatmål

Prosjektet har hatt en tredelt målsetting:

- Å kartlegge hva som finnes av forskningsdata hos den aktuelle gruppa av forskere, og utarbeide en plan for langsiktig forvaltning av deres forskningsdata i nært samarbeid med fakultetene og instituttene, for flest mulig av seniorforskerne.
- Få flytta data fra sårbare lagringsmedier og over til en trygg og langsiktig lagring.
  Bidra, i samarbeid med den enkelte forsker, til at datagrunnlaget beskrives (metadata) og tilrettelegges for fremtidig åpen tilgang og publisering, der det er aktuelt.

Prosjektorganisering

Overordnede prosjektledere har vært Stein Høydalsvik (tidligere faggruppeleder for forsknings- og publiseringsstøtte ved UB) og nåværende faggruppeleder Per Pippin Aspaas. Samtidig har førstebibliotekar Lars Figenschou hatt det faglige (og daglige) prosjektlederansvaret. I tillegg har Abdurhman Kelil Ali vært ansatt som prosjektmedarbeider i 50 % stilling.

Dette prosjektteamet har jobbet spesielt mot utvalgte fagmiljøer ved UiT, samtidig som UBs ressurser og infrastruktur innenfor forskningsdata har inngått som støttefunksjoner for prosjektet – sammen med teknisk støtte og lagringstjenester fra ITA.

Styringsgruppe

Prosjektet har hatt 3 ulike styringsgrupper for hhv 2017, 2018 og 2019. Det vil si at dekan/prodekan/adm. leder har blitt byttet ut årlig, avhengig av hvilke fakulteter og enheter prosjektet har jobbet opp mot det aktuelle året.

Følgende personer fra enhetene har rullert:

- 2017: Prof. Rob Barrett (TMU) og Terje Aspen (fakultetsdirektør, BFE)
- 2018: Prof. Camilla Brekke (NT-fak.) og Prof. Marie-Theres Federhofer (HSL)
- 2019: Prof. Johanna Ericson Sollid (Helsefak.)

Styringsgruppen har (utenom de rullerende personene) bestått av følgende faste medlemmer:

- bibliotekdirektør Johanne Raade (Universitetsdirektørens observatør i styringsgruppen)
- overordnet prosjektleder, hhv Stein Høydalsvik og Per Pippin Aspaas
- prosjektleder (fagansvarlig) Lars Figenschou

Styringsgruppens rolle har primært vært å:

- sikre et godt samarbeid mellom de involverte enhetene og Seniorprosjektet
- påse at prosjektet leverer i samsvar med formål og mål
- kvalitetssikre prosjektet og sørge for at det blir gjennomført innenfor de godkjente ressursrammene

Arbeidspakker og leveranser

Prosjektets gjennomføring var avhengig av at samarbeidet med de utvalgte fakultetene fungerte godt – og det har det gjort. Det måtte skapes en interesse for dette arbeidet både hos fakultetsledelsen på de berørte enhetene, og blant seniorforskerne selv.

Tidsplanen for de prioriterte enhetene har vært:

- 2017: BFE-fakultetet og TMU
- 2018: HSL-fakultetet, NT-fakultetet og IVT-fakultetet
- 2019: Det helsevitenskapelige fakultet

De utvalgte fakultetene og enhetene har fått følgende leveranser (arbeidspakker) levert:

- Avklarings- og strategimøter med institutt- og fakultetsledelsen.
- Kartleggingsmøter. Hva finnes av forskningsdata på de ulike enhetene? Hva bør vi prioritere å ta vare på?
- Informasjons-/opplæringsmøter for grupper av berørte forskere.
- Detaljerte kartleggingsmøter med enkeltforskerne.
- En vurdering og beskrivelse av tilstanden på datasettene (evt. mangel på struktur, sårbarhet, format, graden av unikhet, etc.).
- Innføring i FAIR-prinsippene (som er et sett av retningslinjer som viser hvordan man kan tilrettelegge for videre bruk av forskningsdata)
- Hjelp med overføring av digitalt materiale (data) til sikker lagringstjeneste (ITA).
- Hjelp med strukturering, dokumentasjon og metadata.
- Hjelp med konvertering til persistente og foretrukne filformater.
- Hjelp til, og diskusjoner rundt, graden av sensitivitet.
- Hjelp med arkivering og tilgjengeliggjøring (publisering).
- Tilbud om tette oppfølgingsmøter for alle enkeltforskere som ønsket dette.

Erfaringer fra Seniorprosjektet

Innen de fleste fagområder er det enighet om at god forvaltning og deling av forskningsdata er helt sentralt for å fremme transparens og kvalitet i forskningen. Vi opplever nå at forskerne ser på ordlydene fra forskningsfinansiørene; «As open as possible, as closed as

necessary» (Horizon 2020) og «Åpen som standard» (NFR) som både viktige og riktige. I tillegg har UiT vært tidlig ute med å etablere en egen policy for forskningsdata, noe som har gitt viktig forankring og drahjelp til prosjektet.

Prosjektet har nådd ut til svært mange forskere og ansatte i forskningsadministrative stillinger. For ledelsen på de ulike enhetene, for forskningsmiljøene og for enkeltforskeren, har nok Seniorprosjektet fungert som en slags katalysator. Bevisstheten rundt forvaltning av forskningsdata er, på alle nivåer, en helt annen i dag enn den var ved prosjektstart. Det betyr ikke at UiT eller de enkelte enhetene nå har en praksis der dette blir ivaretatt. Men katalysatoreffekten har vært viktig siden veien kan være veldig lang fra det øyeblikk en forsker bestemmer seg for at det å ta vare på og dele data er en god ide – til tidspunktet hvor det faktisk kan publiseres ett nytt datasett i UiT Open Research Data-arkivet (se kronikk).

At det har vært en slik «treghet» i systemet kan sikkert ha mange grunner. For mange forskere har det frem til nå ikke vært fokus på å tenke struktur, dokumentasjon, beskrivelser og metadata i tilknytning til forskningsdata. Dette har krevd en faglig ryddejobb før vi i prosjektet har kunnet arbeide videre med datasettene i samarbeid med forskeren. Det skal imidlertid også nevnes at det er eksempler på det motsatte – altså at forskerne har hatt meget god orden og at de i sin daglige arbeidsflyt alltid har behandlet sine data på en tilfredsstillende og FAIR måte.

Forskningsdatahåndtering – prinsipper og retningslinjer ved UiT

Det å arkivere, publisere og dele sine data har en veldig sterk smitteeffekt forskere imellom. I løpet av de tre årene prosjektet har vært i gang har det vært en positiv og god utvikling når det gjelder forståelsen av hvor viktig forskningsdataforvaltningen egentlig er ved UiT. Interessen hos forskerne for forskningsdatahåndtering nå i 2020 er på et helt annet nivå enn den var ved prosjektstart i 2017.

Det har altså skjedd veldig mye positivt; en holdningsendring ved alle enheter og på alle nivåer, og særlig med hensyn til forskernes tilnærminger til logikken og argumentasjonen som ligger til grunn for «Open Science» generelt.

UiTs forskningsdatapolicy, «Prinsipper og retningslinjer for forvaltning av forskningsdata ved UiT» (2017) har i så henseende vært svært viktig. Seniorprosjektets oppstart sammenfalt i tid med publiseringen av denne policyen og dette har gitt prosjektet både legitimitet og masse fremdrift. Når det er sagt, har vi også registrert at for mange seniorforskere er dette med full åpenhet (arkivering, publisering og deling av data) rundt egne data en modningssak. En del av seniorforskerne trenger litt tid for å tenke igjennom argumentene som ligger til grunn for UiTs relativt nye forsknings-datapolicy. Når vi så gir dem dette albuerommet, er faktisk samtlige

enige om at deling og publisering av forskningsdata er fornuftig på alle måter (flere siteringer, økt synlighet, innovasjoner går fortere, med mere) både for UiT Norges arktiske universitet som institusjon, og for forskeren selv.

Oppfølging og forslag for å forbedre forvaltningen av forskningsdata ved UiT

Vi bør øke tryggheten rundt forvaltningen av forskningsdata på alle nivåer. For å unngå at UiT gjør seg avhengig av enkeltpersoner og/eller skippertaksmentalitet bør vi normalisere og systematisere arbeidet med forvaltningen av forskningsdata. Dette vil være viktig med tanke på UiTs eierskap og for alle berørte UiT-forskere og miljøene rundt disse.

Dersom UiT ønsker å (fortsatt) inneha en nasjonalt ledende posisjon innenfor Open Science generelt, og innenfor Open Data spesielt, må arbeidet med forvaltningen av forskningsdata implementeres, normaliseres og systematiseres ved alle enheter. Det skal ikke være opp til enkeltforskeren å bestemme hvilke data som skal ivaretas og deles ved UiT. Her må hver enkelt enhet ta ansvar. Enhetene må sørge for at forskerne oppfyller de nasjonale retningslinjene og finansiørenes policyer, og at UiTs egne prinsipper og retningslinjer blir fulgt.

Seks forslag til oppfølging for enhetene ved UiT:

1. Medarbeidersamtalen – utnytt mulighetene

UiT må tydeliggjøre hvilke forventninger de har til sine ansatte som har eller bør ha en funksjon med hensyn til forvaltning av forskningsdata. Samtidig må enkeltforskeren, og særlig de eldre, få informasjon om at han/hun kan få hjelp med forskningsdata-håndteringen. De som har ansvaret for medarbeidersamtalen bør vise til forskningsdataportalen og oppfordre forskerne til å utnytte det store utvalget av opplæringsmuligheter innenfor forskningsdata-håndtering ved UiT.

Det er videre helt essensielt at forskerne er informert om Eierskap til data, plikter og rettigheter og forskerens ansvar og Retningslinjer for oppfølging av datahåndteringsplaner ved UiT. Alle nye forskningsprosjekter må også hensynta, og jobbe etter FAIR-prinsippene.

2. Følg opp de kartleggingsjobbene som allerede er utført

Mange enheter har nå god oversikt over hva som finnes av data på enheten. Ta diskusjonen om hva som bør tas vare på for fremtiden og sørg for å få flyttet eksisterende data fra sårbare lagringsmedier (minnepenner, kontormaskiner, felt/loggbøker) over til UiT-godkjente systemer for trygg og langsiktig lagring.

3. Følg opp arbeidet med datahåndteringsplanene

Fakultetene/enhetene har ansvar for oppfølging og godkjenning av datahåndteringsplaner og innpasse dette i rutiner for sine forskningsprosjekter (se Retningslinjer for oppfølging av data-håndteringsplaner ved UiTs enheter).

4. Hvor publiserer og deler forskerne sine forskningsdata?

Alle institutter/enheter bør innhente oversikter over dette. Dokumentasjon kommer til å bli etterspurt – kanskje særlig fra finansiørene. Vi anbefaler at alle vitenskapelig ansatte skaffer seg en ORCID-identifisering som brukes til å identifisere forskere. Det vil gjøre det enkelt for enhetene å fremskaffe en slik oversikt.

5. Tilby støtte og ressurser

Allerede eksisterende data bør dokumenteres, struktureres, arkiveres og publiseres. Vurder å ansette en (tidsbegrenset) internship og/eller data-steward. Hva med å benytte ph.d-studenter eller postdoktorer? Dersom man kjenner konteksten rundt dataene kan man få gjort mye i løpet av kort tid (2–6 mnd). En del seniorer har f.eks. tidsserier liggende – disse kan det være viktig å få publisert og delt.

6. Utnytt kompetansen som allerede befinner seg ved UiT

Ved UB finnes det kompetanse på forvaltning av forskningsdata. UB sørger for at dataene er så FAIR som mulig, og på den måten tilføres ofte datasettene en ytterligere verdi. UB tilbyr blant annet opplæring og undervisning i forskningsdatahåndtering, hjelp med arkivering, publisering og deling, og hjelp med kuratering av de enkelte datasett.

Konklusjon

Forslagene over vil alle bidra til at fremtidig forskning kan dra nytte av, og bygge videre på, tidligere fremskaffa forskningsdata. Samtidig oppfyller vi forventninger og krav fra Kunnskapsdepartementet, fra finansiørene, og fra vår egen institusjon.

God forvaltning av forskningsdata vil utvilsomt gi nye perspektiver for fremtidige forskningsprosjekter, i tillegg til at det er veldig kostnadsbesparende over tid. Datainnsamling og datagenerering er veldig kostbart. Å legge til rette for gjenbruk kan derfor bety store besparelser.

Sluttelig vil selvsagt økt publisering av forskningsdata med utspring fra UiT også medføre økt synlighet, både for UiT som institusjon, for de ulike enhetene, og for enkeltforskerne.

Seniorprosjektet i tall og filer

Total har vi hatt i overkant av 120 ansikt-til-ansikt møter med FSU (forskningsstrategisk utvalg), instituttledermøter, dekaner, forskningsgruppeledere, forskningsgrupper og enkeltforskere. På de

ulike enhetene har prosjektet hatt e-postkontakt og/eller personlig kontakt med:

- Fakultet for biovitenskap, fiskeri og økonomi UiT: 45 forskere
- Norges arktiske universitetsmuseum: 9 forskere
- Handelshøgskolen ved UiT: 4 forskere
- Fakultet for naturvitenskap og teknologi: 12 forskere
- Fakultet for humaniora, samfunnsvitenskap og lærer-utdanning: 28 forskere
- Fakultet for ingeniørvitenskap og teknologi: 3 forskere
- Det helsevitenskapelige fakultet: 35 forskere

Seniorprosjektet har fungert som fødselshjelp for mange (eldre) datasett. Det vil si at dataene har blitt flyttet til sikre lagringsmedier, dokumentert, arkivert, publisert og dermed delt i et åpent arkiv, og fortrinnsvis i UiT Open Research Data-arkivet.

Når det gjelder hvor mange datasett og hvor mange filer det dreier seg om, så er det litt vanskelig å tallfeste dette siden prosjekt-medlemmene samtidig har vært en del av Forskningsdatagruppen ved UB, som til daglig forvalter driften av UiT Open Research Data-arkivet ved UiT. Det er derfor en litt flytende grense mellom hvilke arkiverte datasett som kan tilskrives Seniordataprosjektet, og hvilke som faller utenom. Per 31.12.2019 hadde UiT Open Research Data 649 publiserte datasett. Seniordataprosjektet har bidratt til flere hundre av disse, med tusenvis av filer.

Seniorprosjektet har (sammen med Forskningsdatagruppen ved UB) fulgt opp både større og mindre forskningsprosjekter, i tillegg til alle enkeltforskerne.
Gode eksempler på variasjonen i dette er:

- NMDC (Norwegian Marine Data Center)
- TGO (Tromsø Geophysical Observatory; Nordlysdata)
- Trivsel i Tromsø (Digital mobbing og psykisk helse hos barn; Steinar Thorvaldsen ILP)
- Spring migrants to Tromsø (Rob Barrett, TMU; tidsserier, trekkfugler)
- Trolldomsprosessene i Finnmark (Rune Blix Hagen, HSL)

Vedlegg

Hvordan håndterer du forskningsdata?

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

DataverseNO: A National, Generic Repository and its Contribution to the Increased FAIRness of Data from the Long Tail of Research

Philipp Conzett

Abstract

Research data repositories play a crucial role in the FAIR (Findable, Accessible, Interoperable, Reusable) ecosystem of digital objects. DataverseNO is a national, generic repository for open research data, primarily from researchers affiliated with Norwegian research organizations. The repository runs on the open-source software Dataverse. This article presents the organization and operation of DataverseNO, and investigates how the repository contributes to the increased FAIRness of small and medium sized research data. Sections 1 to 3 present background information about the FAIR Data Principles (section 1), how FAIR may be turned into reality (section 2), and what these principles and recommendations imply for data from the so-called long tail of research, i.e. small and medium-sized datasets that are often heterogenous in nature and hard to standardize (section 3). Section 4 gives an overview of the key organizational features of DataverseNO, followed by an evaluation of how well DataverseNO and the repository application Dataverse as such support the FAIR Data Principles (section 5). Section 6 discusses how sustainable and trustworthy the repository is. The article is rounded up in section 7 by a brief summary including a look into the future of the repository.

1. The FAIR Data Principles

Data constitute the core assets within many scientific disciplines. New knowledge and insight are often drawn from the analysis of data. However, in traditional scholarly communication, research data have not been granted the attention one would expect given their importance for the advancement of science. Among the different kinds of output from research activities, describing research results in articles and books published in recognized venues is still the most rewarding way of communication for the vast majority or researchers. Data, on the other hand, are – with the exception of some fields – rarely published and shared. This situation is quite surprising considering that in most cases scientific results and claims cannot be verified without access to the underlying data. There are also other unfortunate consequences of data not being made accessible, among

https://doi.org/10.7557/15.5514

others the fact that data cannot be reused by other researchers, which potentially results in duplication of efforts; and that researchers may miss important possibilities for new collaborations arising from shared data.

Research funders and other stakeholders have for quite some time been aware of the unfortunate consequences of research data not being reusable, and the urge to improve reusability of data has been continuously growing as research methods and tools have become increasingly digital. As one of the influential international stakeholders, the Organisation for Economic Co-operation and Development (OECD) addressed the problem of reduced reusability as early as in 2004, when they adopted a declaration on access to research data from public funding. This declaration resulted in the OECD Principles and Guidelines for Access to Research Data from Public Funding, which was published in 2007 (OECD, 2007). Similar guidelines and recommendations have since been adopted by other funding agencies and research organizations (Christian et al., 2020; Crosas et al., 2018; Neylon, 2017). Drawing on this early work, one of the most influential papers in the field of data management saw the light of day in 2016, when a diverse group of stakeholders representing academia, industry, funding agencies, and scholarly publishers postulated a set of principles on how to improve infrastructure supporting the reuse of research data. They referred to these principles as the FAIR Data Principles (Wilkinson et al., 2016).

FAIR data are data that are Findable, Accessible, Interoperable, and Reusable. A key feature that according to the authors distinguishes the FAIR Data Principles from similar initiatives is their emphasis on machine-actionability, meaning that not only humans should be able to find, access and reuse research data, but also machines. While the results from data analysis ultimately are to be understood and interpreted by humans, machine-actionable research data can increase the efficiency of data management and analysis considerably. Also, due to the nature and/or amount of data in some scientific fields, machine-actionability is not only a question of effectiveness, but simply a necessity. In their seminal article, Wilkinson et al. (2016, p. 4) summarize the FAIR Data Principles as follows:

> *To be Findable:*
> F1. (meta)data are assigned a globally unique and persistent identifier.
> F2. data are described with rich metadata (defined by R1 below).
> F3. metadata clearly and explicitly include the identifier of the data it describes.
> F4. (meta)data are registered or indexed in a searchable resource.
>
> *To be Accessible:*
> A1. (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2. metadata are accessible, even when the data are no longer available.

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with detailed provenance.
R1.3. (meta)data meet domain-relevant community standards.

The implication of the different parts of the FAIR Data Principles will become evident from the detailed review presented in section 5.

As summarized in Figure 1, FAIR research data may be considered as both the output of and the input to good data stewardship throughout the entire lifecycle of research data. Let us start with the processes illustrated with the orange arrows. Early in a research project, data are collected or generated, and possibly processed (e.g. annotated or enriched in other ways). During this active phase of the research project, it is essential to have good routines in place for organizing and describing the data as well as for data storage. Processed data are then analysed, and the findings from data analysis are usually presented in articles or books. At the same time, the background data for these publications should be archived and shared with the scientific community and the greater public. An important step before data sharing is to document the data to enable other researchers to understand and reuse them. The grey box at the top of the figure indicates that all these activities should be carried out in line with policies, standards and good practice recommendations for research data management, and that good planning lays the foundation for successful data management. However, contrary to widespread belief, planning is not a one-off task to get done with at the outset of a research project. Rather, a data management plan (DMP) only reveals its full potential when used as an active document that is updated and revised throughout the project. In Figure 1, this is indicated by the multiple grey arrows below the grey box. The ideal output of the research activities described so far are FAIR research data, illustrated in the green arrow-box.

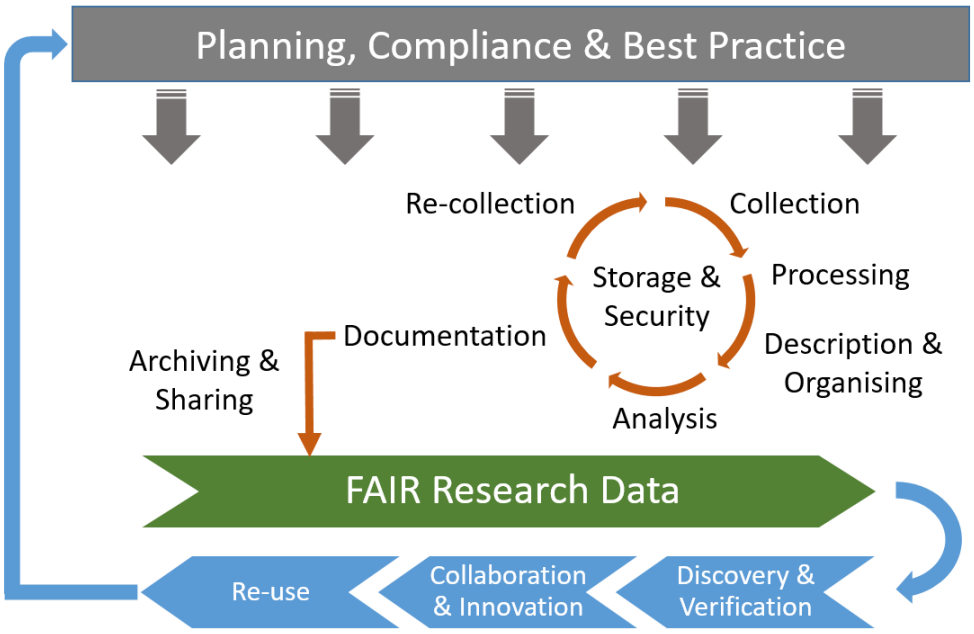


Figure 1: The FAIR Research Data Lifecycle.

These data may, in turn, serve as input to other projects, illustrated by the blue arrows and arrow-boxes. FAIR research data can be discovered, and they may be used to verify the scientific findings and claims that are set forth based on them. Discoverable data is a driving force for new collaborations between researchers as well as innovation within the scientific community, but also in society more broadly. In more general terms, FAIR research data may be reused, which brings us to the beginning of the next round of the lifecycle of research data. Considering potential reuse of data should be a natural part of the planning of any research project in need for empirical evidence.

2. How to turn FAIR into reality?

The FAIR Data Principles are a set of general guidelines, they do not include specific advice on how these principles may or should be implemented. Many research support services provide guidance for researchers and others involved in research data management on how to organize and document their data to make them reusable. There are also numerous high-level initiatives working for the uptake of the FAIR Data Principles on a more global scale. The most prominent example in Europe is probably the establishment of the European Open Science Cloud (EOSC) (European Commission, n.d.).

As has been pointed out by several experts, FAIR is not a binary concept. Instead, data – or more broadly: digital objects – may be more or less FAIR, and we can thus speak of degrees of FAIRness. Recommendations on FAIR data management are thus meant to support a transition towards increased FAIRness of research data at large.

Ultimately, the transition towards increased FAIRness requires change in research culture. Brian Nosek uses two theories to illustrate the mechanisms and progress involved in research culture change. The first one is the Theory of Diffusion of Innovation, proposed by Everett Rogers in 1962 (Rogers, 2003). Nosek suggests that the normal distribution graph used in Rogers' theory and redrawn in Figure 2 also applies to the spread of change in research culture (Nosek, n.d.).

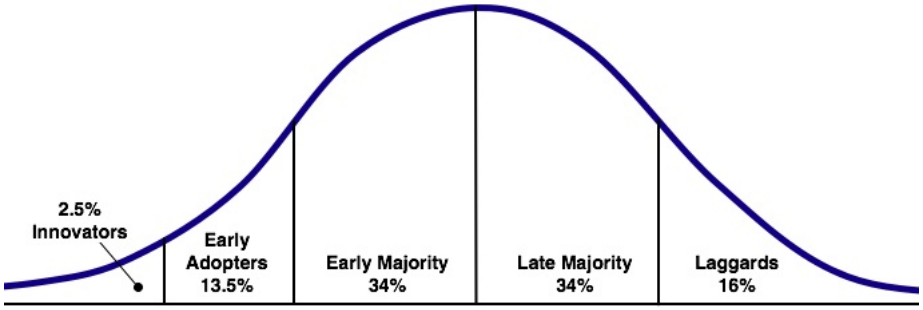


Figure 2: The Diffusion of Innovation. Redrawn from Rogers (2003, p. 281).

According to this view, a change or innovation in research culture is first initiated and spearheaded by a small group of innovators, followed by a somewhat larger group of early adopters, before the change is adopted by an early majority and afterwards by a late majority of members of the research community. Finally, the change spreads to the last group of adopters, called laggards in Rogers' theory.

The second theory adopted by Nosek is the Theory of Human Motivation, originally proposed in 1943 by Abraham Maslow (Maslow, 1943). Maslow postulates a hierarchy of needs to illustrate how human behaviour is governed by motivation. Adapting Maslow's hierarchy of needs to the realm of research, Nosek uses the pyramid in Figure 3 to illustrate the different motivational factors and driving forces behind cultural change. In this view, basic infrastructure including tools and skills are necessary to make change in research culture possible. Turning this infrastructure more user-friendly makes it easy for members of the research community to adopt new practice. Once new practice has spread to and is recommended by (a large part of) a research community, its adoption may be considered normative. As a further step in advancing the uptake of new practice, incentives may be introduced that make the adoption of the practice rewarding. Finally, the implementation of new practice may be made required by policies.

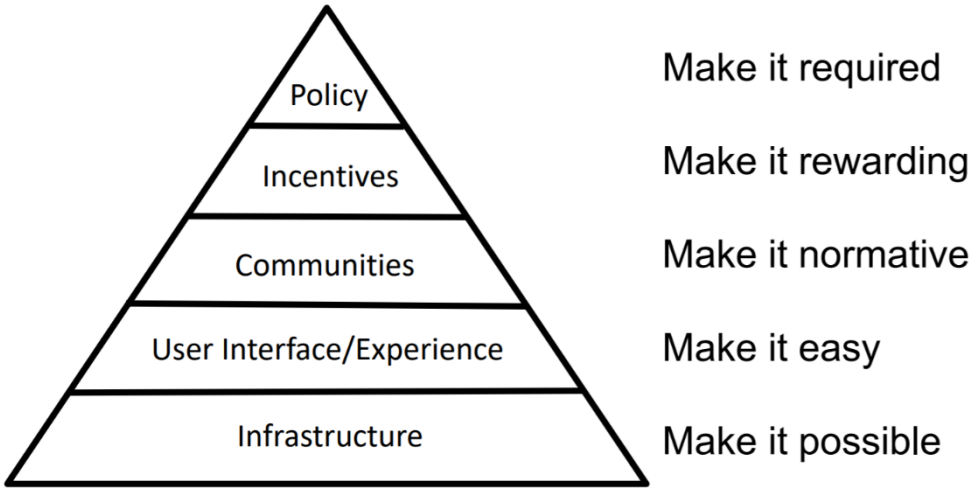


Figure 3: Motivational factors and driving forces behind change of research culture. From Nosek (n.d.), licensed under CC0 1.0 Universal.

In Figure 4, Nosek combines his adaption of Rogers' Theory of Diffusion of Innovation and Maslow's hierarchy of need. In this view, basic infrastructure making the change of research culture possible is sufficient motivation for the small group of innovators to adopt new practice. Early adopters go along with the innovators once adoption has been made easy. Adoption by the early majority is mainly driven by the new practice becoming part of community norm, while rewarding incentives are the main motivation needed for the late majority to change their behaviour. Finally, policy requirements seem to be the last resort to motivate the group of laggards to comply with what by then probably is recognized as a de facto standard in the research community.

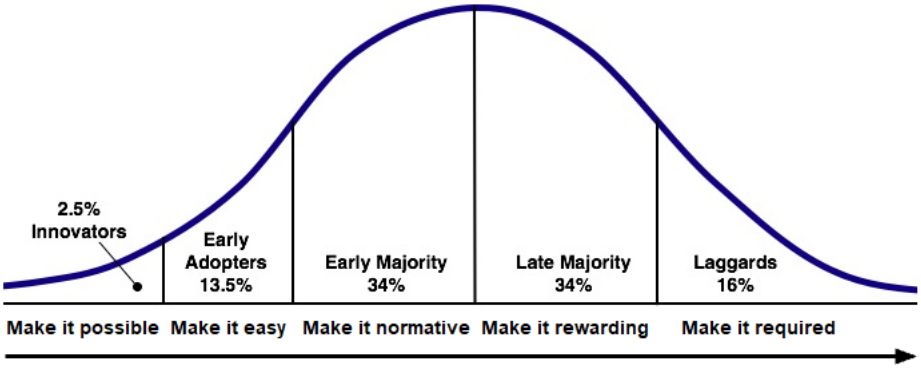


Figure 4: Change in research culture: diffusion and motivation. Slightly adapted from Nosek (n.d.), licensed under CC0 1.0 Universal.

To my knowledge, Nosek's view of how change in research culture diffuses in research communities has not been verified by empirical data. However, applied to the process of increasing the FAIRness of research data, there are some observations I have made in my work with providing support services for research data management for the last six years which to some degree substantiate Nosek's view.

First, there are parts of research communities that started adopting the FAIR Data Principles without there being other resources in place

than the basic infrastructure making the adoption possible. Parts of the bioinformatics community may here serve as an example.

Second, for quite some time researchers have had access to data repositories and other support services making it rather easy to make their data at least partly FAIR. Nosek suggests that – seen from a post-hoc perspective once the culture change has been accomplished – the group of early adopters amounts to 13.5% of the research community. To get a rough indication of how large the group of early adopters of the FAIR Data Principles currently might be in Norway, one could compare the number of unique authors of published data-sets with the number of unique authors of publications of research results in anthology chapters, articles and monographs (books). In a small case study, I obtained these numbers limited to researchers affiliated with my own university, UiT The Arctic University of Norway (UiT), and limited to outputs published in 2019. The background data for this small investigation including a description of the methods and tools used to obtain these numbers are available in Conzett (2020). There were 20 unique UiT-affiliated researchers who published one or more datasets in 2019. In the same year, there were 1736 unique UiT-affiliated researchers who published research results in recognized publishing venues. The group of UiT-affiliated researchers who published data in 2019 may thus be said to represent 1.15% of all UiT-affiliated researchers who published research results in the same year. These numbers are based on a very small selection of researchers and limited to the span of only one year. I still argue that they can give us a rough indication. The results from this small investigation suggest that the percentage of researchers making use of easily accessible resources to make data more FAIR is still much lower than the 13.5% indicated by Nosek for the group of early adopters.

Third, there seems to be considerable agreement among research data professionals that strong support and recommendations from within the research community as well as rewarding incentives are crucial to advance the further uptake of the FAIR Data Principles. On the other hand, the effectiveness of policy requirements may vary somewhat. Some research funders and journals have introduced policies that require researchers to make their data openly available. Such policies are most effective when non-compliance results in direct negative consequences like repayment of funding or the pre-clusion of article or book manuscripts from being accepted for review. Journals' data policies are thus an effective incentive for researchers to make their data available. Currently, most universities do not reinforce the compliance with their research data policies. However, institutional policies are still important as they help provide useful legitimation for institutional support services (see for example Figenschou in this volume).

In addition to the driving forces described above also other factors affect the way in which the FAIR Data Principles may be turned into

reality. The properties of research data themselves is one such factor, which is the topic of the next section.

3. Data from the long tail of research

Research data come in different forms, and this diversity may pose some challenges to responsible data stewardship. An overall – though somewhat simplified – distinction commonly made is the one between big data and small data (Borgman, 2015, pp. 8–10). The distribution of scientific work along the continuum between big and small data has been described as a head with a long tail, meaning that a small number of research projects deal with very large volumes of data, whereas the vast majority of researchers work with medium-sized or small volume data. The distribution is illustrated in Figure 5.

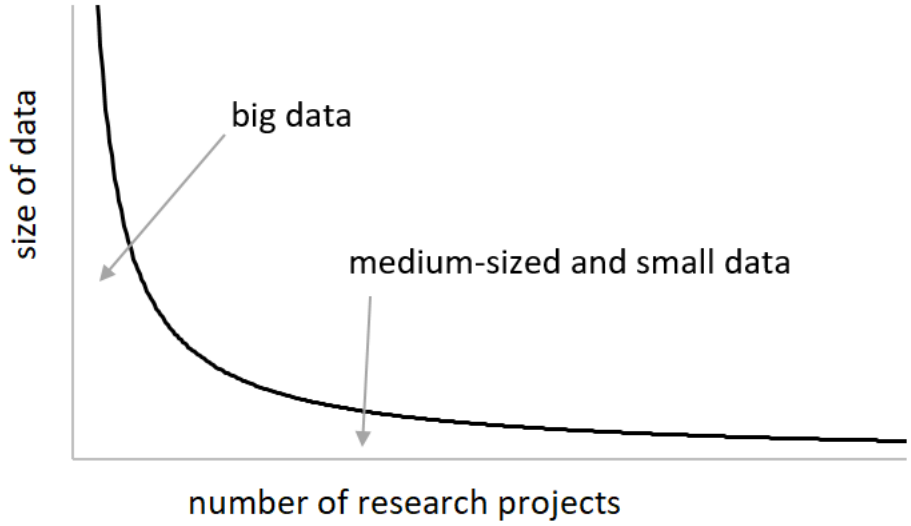


Figure 5: Distribution of research projects according to data size.

On a generalised level, the distribution shown in Figure 5 is also said to reflect properties of research data other than data volume or size (Borgman, 2015, pp. 8–10; Heidorn, 2008; The e-IRG Task Force on the Long Tail of Data, 2016). Big data tend to be homogeneous in content and form, whereas small datasets often represent a multitude of data types. Big data are often handled within large projects, which makes it more urgent to agree on common standards for data structure, file formats, documentation etc. at an early stage of the project. In such projects, it is also common to share infrastructure, tools and services like data curation. In small or less data-driven fields and/or in small or medium-sized projects, researchers can more easily adapt data management to the problems they are set to solve, without putting the progress of the project at risk. "The down-side to such flexibility is" according to Borgman (2015, p. 10) "the lack of standards on which to base shared infrastructure and the lack of critical mass to develop and sustain shared data resources". In other words, the further down the tail of the distribution in Figure 5 a research project may be placed, the more diverse the data are in content, structure and representation, and the less common are

shared standards and resources to support uniform data management (Borgman, 2015, p. 10).

With the rapid development of computing capacity and technology, big data have received much attention from research funders and from society at large during the last decade or so. However, as emphasized by Borgman (2015, pp. 8–9), "only a few fields, such as astronomy, physics, and genomics in the sciences; macroeconomics in the social sciences; and some areas of digital humanities; work with very large volumes of data in an absolute sense". Although the enormous value of small and medium-sized data is recognised by advocates of Open Science and cross-disciplinary research, the e-IRG Task Force on the Long Tail of Data (2016, p. 4) argues that the importance of long-tail data had become out of sight of funders and policy makers with the advent of Big Data. This view is supported in a 2018 report from the European Commission expert group on FAIR data: "The so-called 'long tail' of research remains poorly catered for, and vast amounts of data produced in research are not FAIR and currently lack long-term stewardship" (European Commission, 2018, p. 55). In the work to remedy this shortcoming, the design principles and organizational models of research data repositories are given a crucial role (The e-IRG Task Force on the Long Tail of Data, 2016, p. 4). Traditionally, a fair amount of the data from the long tail of research that are made available, find their home in general or generic research data repositories, one of them being DataverseNO.

4. What is DataverseNO, and how does it work?

DataverseNO (re3data.org, 2017) is a national, generic repository for open research data. The repository is owned and operated by UiT The Arctic University of Norway. The technical infrastructure of the repository is based on the open source application Dataverse (n.d.), which is developed by an international developer and user community led by Harvard University. DataverseNO supports the FAIR Data Principles and has recently been certified as a sustainable and trustworthy repository by the CoreTrustSeal (n.d.).

Established in 2017, DataverseNO has its origin in UiT's institutional research data repository, UiT Open Research Data (n.d.), which was launched in 2016. Before that, UiT had been developing and running support services for research data management for a couple of years (Conzett & Østvand, 2018). Both UiT Open Research Data and DataverseNO have grown out of the Tromsø Repository of Language and Linguistics (re3data.org, 2015), the first research data repository service established at UiT. The Tromsø Repository of Language and Linguistics (TROLLing) is a domain-specific repository that was initiated back in 2013 by linguists at UiT and has since its launch in 2014 been an open and free repository where linguists worldwide can deposit and publish their research data (Conzett, 2019; GÉANT & UNINETT, 2019). TROLLing as well as UiT Open Research Data are now part of DataverseNO constituting each its own collection.

All datasets deposited into DataverseNO are curated by research data support staff before they are published. Published data can be edited. Any changes are subject to a new curatorial round and result, on publication, in a new version of the dataset with all previously published versions still being available.

Figure 6 gives an overall outline of the organization of DataverseNO. The repository structure of DataverseNO is outlined in the middle box in the chart. Norwegian research organizations may enter into a partner agreement with UiT to use DataverseNO as an institutional repository for open research data. Datasets from researchers affiliated with DataverseNO partner institutions are published in designated institutional collections, of which there are currently nine. Individual researchers from Norwegian research organizations that are not partnering with DataverseNO can publish their data in the top-level collection of the repository (indicated with Dataset 10, 11, … in the repository structure box in Figure 6). Another type of collection within DataverseNO are special collections, which may be project-based and/or subject-based. Such collections may be open for contribution from researchers outside Norwegian organizations. TROLLing is a thematic collection, and currently the only special collection in DataverseNO. All special collections within DataverseNO are at the full responsibility of a DataverseNO partner institution.

As indicated in the top box in the DataverseNO Organization Chart, a set of organizational documents regulate the organization of the repository, including its structure, governance, data curation, and its Designated Community. At the core of these documents we find the DataverseNO Policy Framework (n.d.), which – in addition to a general part introducing the scope and mission of the repository as well as defining some core concepts – consists of four policies, namely, the DataverseNO Access and Use Policy, the DataverseNO Accession Policy, the DataverseNO Deposit Agreement, and the DataverseNO Preservation Policy.

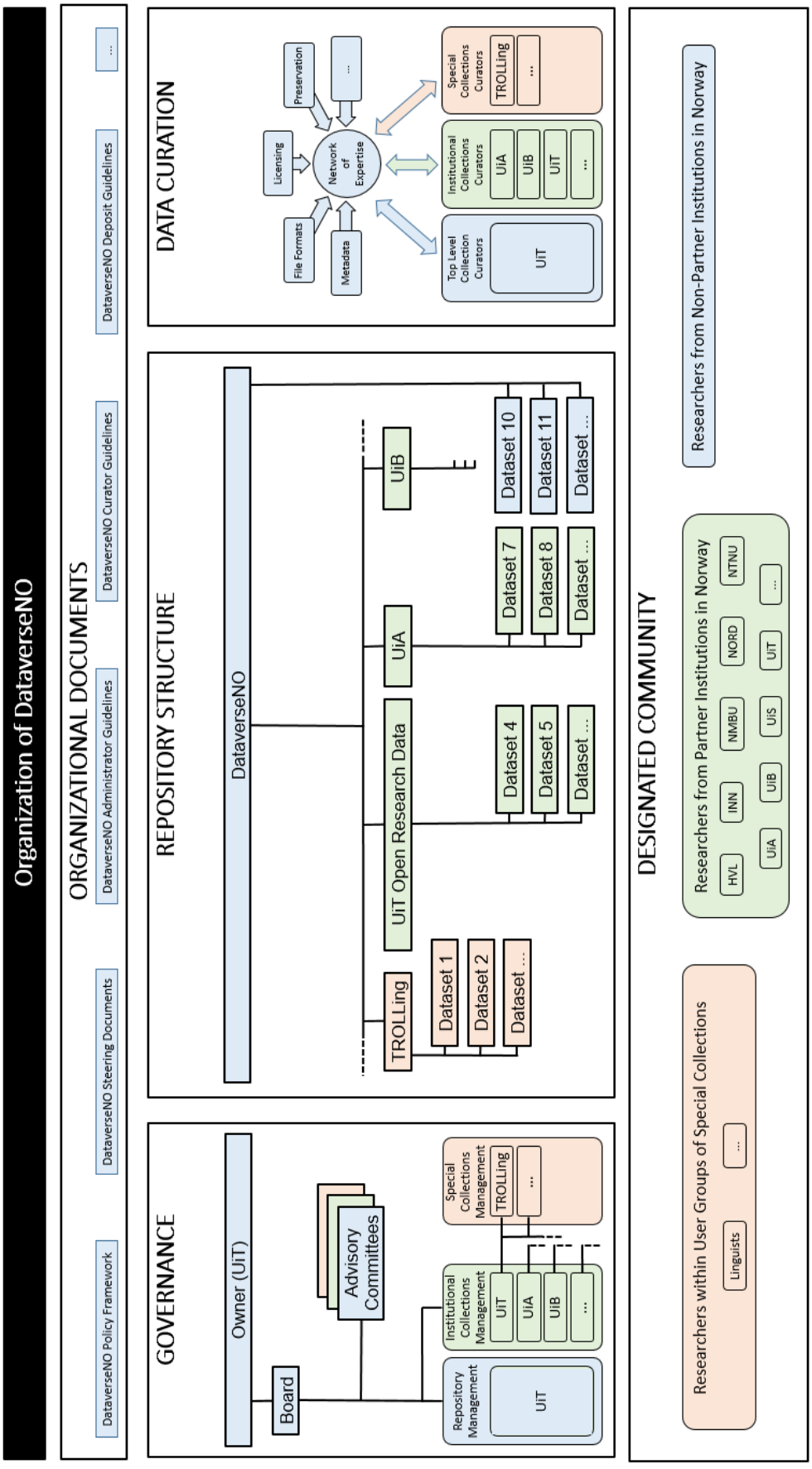


Figure 6: DataverseNO Organization Chart. From Organization of DataverseNO (n.d.), licensed under CC0 1.0 Universal.

The DataverseNO Access and Use Policy outlines DataverseNO's commitment to facilitating maximum access and use of research data published in the repository, and describes the mechanisms DataverseNO implements to fulfil this commitment, including the provision of persistent URLs and the assignment of Digital Object Identifiers (DOIs) for reliable discovery and citing of published Datasets, and facilitating the indexing of descriptive metadata by search engines. *The DataverseNO Accession Policy* explains what DataverseNO can accept for publication, which is essentially research data that can be made openly available, are appropriately documented, and are stored in preferred file formats. With the exception of special collections like TROLLing, DataverseNO accepts only data from researchers affiliated with Norwegian research organizations. At initial sign up, depositors are required to accept the *DataverseNO Deposit Agreement*, which gives DataverseNO a non-exclusive right to distribute published data on the internet, and the right to copy and convert published files to new file formats to the extent required to ensure secure storage and sustainable long-term preservation. *The DataverseNO Preservation Policy* describes DataverseNO's commitments and approaches to responsible and sustainable stewardship of published datasets in the long term, including preservation strategies and levels of preservation. The concrete preservation actions are specified in the DataverseNO Preservation Plan.

To implement the commitments stated in the DataverseNO Policy Framework, the DataverseNO Guidelines provide concrete guidance for the three main stakeholder categories involved in the operation of the repository. The DataverseNO Deposit Guidelines (n.d.) describe how depositors are required to prepare their data prior to deposit, and how to create and get datasets published in DataverseNO. The DataverseNO Curator Guidelines (n.d.) contain documentation about how curators are expected to review deposited data before publication. The DataverseNO Administrator Guidelines describe the main tasks to be carried out by repository as well as collection administrators, including user and access control management.

The DataverseNO Partner Agreement regulates the responsibilities of UiT as the owner and operator of the repository on the one hand, and of the partner institutions on the other hand. Most importantly, the agreement states that all DataverseNO policies and guidelines apply to the entire repository with all its collections. Partner institutions are thus obligated to manage their institutional collections in compliance with the DataverseNO policy framework.

Data curation is carried out by data curators at the DataverseNO partner institutions, as indicated in the middle right box in Figure 6. Datasets deposited in the top-level collection are curated by data curators at UiT. Data curators are staff members employed at the partner institutions, usually at the library of the institution. Data curators are responsible for ensuring that data published in each collection within DataverseNO (including the top-level collection) are

curated according to the DataverseNO policies and guidelines, and in line with best practice recommendations and the needs of the different user communities at stake. Curators communicate with the different user communities represented in the collection(s) they curate, e.g. during curation, but also through other channels and in other venues. Curators also communicate with the management of their collection, and with curators of other collections within DataverseNO through the DataverseNO Network of Expertise. This network of curators covers the different aspects of data curation, including metadata and documentation, file organizing and file formats, and licensing. In addition to enabling knowledge and experience exchange, this network also makes sure that curation practices across the repository are aligned with the DataverseNO policies and guidelines. The network also seeks to align curation practices across institutional collections from different partner institutions containing data from the same or similar scholarly disciplines.

The governance of DataverseNO is illustrated in the middle left box in Figure 6. As the owner and operator of DataverseNO, UiT offers the repository as a service to other research organizations and to individual researchers from research organizations in Norway.

The Board of DataverseNO has the overall responsibility for DataverseNO, with a mandate provided by the university management of UiT.

Collections within DataverseNO may have their own advisory committees which give advice to the collection managers as well as to the Board of DataverseNO on high-level aspects of the operation and development of a specific collection as well as the entire repository. The Designated Community may raise high-level or general issues with representatives from the advisory committee of the collection at stake. Currently, only TROLLing has formally established an advisory committee, the TROLLing Scientific Advisory Board, who provide their advice to the TROLLing managers.

The operation of institutional collections is embedded in the research support services and endorsed by the institutional management at the DataverseNO partner institutions. At each partner institution, there are procedures and venues in place where research support units, such as the university library, discuss issues with representatives from the different research communities at the institution. Feedback from such discussions is provided to the managers of the institutional collections. On their part, managers for institutional collections discuss advice and feedback from the user groups of their institutional collections in the Advisory Committee for DataverseNO. This committee, illustrated with the blue box in the middle of the GOVERNANCE section of the DataverseNO Organization Chart, consists of representatives from all DataverseNO partner institutions (usually the collection managers), and the managers of DataverseNO. The members of the DataverseNO Advisory Committee meet at least

twice a year to discuss issues concerning the organization of DataverseNO, including governance, policies and guidelines, repository structure and operation (including functionality), data curation, and issues raised by the Designated Community. Requests and advice from the DataverseNO Advisory Committee are communicated to the Board of DataverseNO and to the managers of the institutional collections by the Repository Management.

The DataverseNO repository is managed by staff from the University Library and the IT department at UiT. They are responsible for the management, maintenance, development and the daily operation of the repository, and they take care of the DataverseNO policies and guidelines, communication with the Board of DataverseNO, communication with and training of collection managers, the operation of the DataverseNO Advisory Committee, the configuration of the repository, establishment and configuration of institutional collections, training of collections managers, user management, the implementation of new functionality and procedures to be used in the repository, preservation planning, and the certification of the repository.

The managers of institutional collections within DataverseNO are responsible for the management and operation of the collection, including compliance of the institutional collection and underlying collections with the DataverseNO policies and guidelines, user management of collection curators, training of and communication with collection curators, establishment and configuration of underlying collections, communication with the DataverseNO repository management, communication with the management and the user groups at the partner institution as well as representing the collection in the DataverseNO Advisory Committee.

The managers of special collections have many of the same responsibilities as institutional collection managers, but limited to the scope of the collection. They communicate with the advisory committee for the collection – if applicable.

The bottom section of Figure 6 illustrates the Designated Community of DataverseNO. Since the repository provides free and open access to its collections, the Designated Community of DataverseNO consists of both data contributors and data users. Data users include primarily researchers and research organizations, but also any other stakeholders in society reliant on access to knowledge, e.g. journalists, teachers, industry as well as the greater public. The interaction between data users and the repository happens primarily through direct contact with the contact person(s) for each dataset, and through the general contact information provided for each collection. In a more narrow sense, the Designated Community of DataverseNO can be described as the different user groups that in addition to being data users also are data contributors to the repository. As outlined in the DataverseNO Organization Chart, these user groups fall into three main categories:

- researchers from Norwegian research institutions that are partners of DataverseNO
- researchers working within the scope of any special collection within the DataverseNO repository
- researchers from Norwegian research institutions that are not partners of DataverseNO

Each collection is organized and managed in a way that ensures that the needs of the user group are met to the largest possible extent.

To round off this section, Table 1 gives an overview of the numbers of published datasets in Norwegian research data repositories. The numbers were retrieved from DataCite Search (n.d.) on April 6, 2020, for all repositories using Digital Object Identifiers (DOIs) provided through Unit – Directorate for ICT and joint services in higher education and research, which is the Norwegian national DOI provider.[1]

Table 1: Published datasets with DOI in Norwegian repositories as of 6 April 2020. Sources: DataCite Search and DataverseNO.[2]

Repository	2014	2015	2016	2017	2018	2019	2020	SUM
BI						1		1
DataverseNO	21	11	31	37	201	371	28	700
NIBIO				1	1			2
NILU				1	3			4
NMDC				8	29	10	7	54
NPOLAR			95	40	10	46	2	193
NSD			200	84	1 387	249	25	1 945
UNINETT	17	32	13	22	28	34	6	152

The numbers presented in Table 1 are illustrated as stacked columns in Figure 7.

[1] The table includes only DOI registrations of the resource type “dataset”. DataverseNO also assigns DOIs at file level. Unfortunately, DataCite does currently not distinguish between dataset DOIs and file DOIs. Therefore, the numbers for DataverseNO were obtained directly from the repository. The numbers for DataverseNO also include datasets that were published in TROLLing and UiT Open Research Data before these repositories were included as collections within DataverseNO in 2017.

[2] Abbreviations: BI = Norwegian Business School; NIBIO = Norwegian Institute of Bioeconomy Research; NILU = NILU – Norwegian Institute for Air Research; NMDC = Norwegian Marine Data Centre; NPOLAR = Norwegian Polar Institute; NSD = NSD - Norwegian Centre for Research Data; UNINETT = UNINETT Sigma

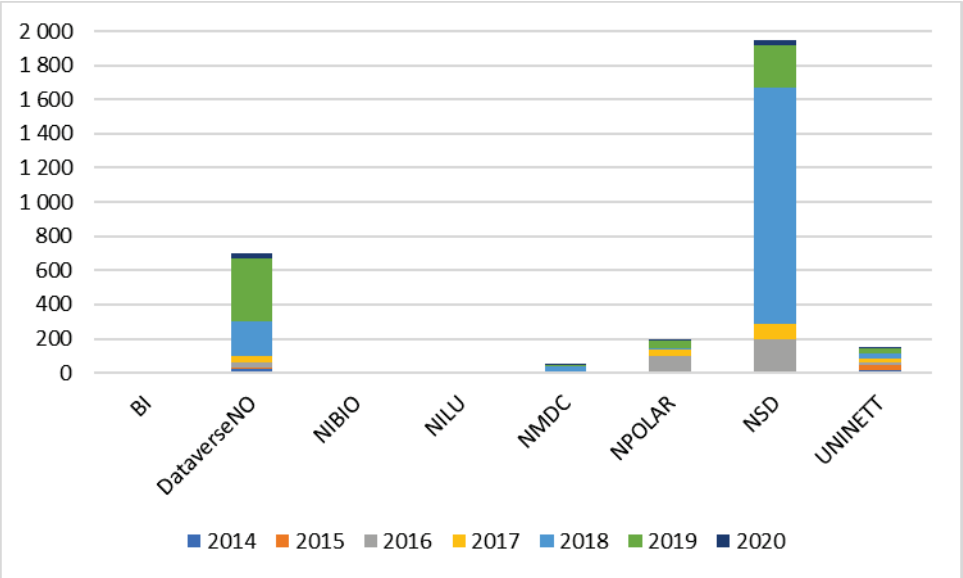


Figure 7: Published datasets with DOI in Norwegian repositories as of 6 April 2020. Sources: DataCite Search and DataverseNO.

Measured in published datasets with DOIs, DataverseNO is thus the second largest research data repository operated in Norway. This position is quite remarkable considering the brief history of the repository.

5. How FAIR are data published in DataverseNO?

Having presented the key organizational features of the DataverseNO repository, and with the overview of the FAIR Data Principles presented in section 1 in mind, let us now have a closer look at how FAIR the data published in DataverseNO are. To answer this question, I will build on a recent systematic overview of how the FAIR Data Principles are implemented in the repository application Dataverse (Crosas, 2020), supplied with information about how these implementations are applied in DataverseNO, and how DataverseNO in addition supports the FAIR Data Principles through deposit requirements and data curation.

To be Findable

Jacobsen et al. (2019, p. 14) summarize the Findable principle in FAIR as follows:

> Digital resources should be easy to find for both humans and computers. Extensive machine-actionable metadata are essential for automatic discovery of relevant datasets and services, and are therefore an essential component of the FAIRification process.

The Findable principle has four elements. According to F1, Findable (meta)data are assigned a globally unique and persistent identifier (PID). Dataverse has implemented F1 by supporting two PID systems, DOI and Handle. All datasets get assigned a PID, whereas PID assignment at file level is an optional feature in Dataverse. DataverseNO

has been using DOI since the repository was established in 2017. The repository has also adopted DOIs at file-level for all files deposited after we in 2018 upgraded the software to version 4.9 where this feature was introduced in Dataverse.

According to F2, Findable data are described with rich metadata. Dataverse supports F2 by providing a discovery metadata schema based on the following widely-used discovery metadata standards in human- and machine-readable formats: Dublin Core (n.d.), Documentation Data Initiative (DDI) (n.d.), DataCite (n.d.), and Schema.org (n.d.). Among the rich options for adding discovery metadata to datasets, DataverseNO in particular emphasizes three mandatory fields: Title, Description, and Keywords. Optionally, customized metadata schemas can easily be configured in Dataverse in addition to the built-in schemas. This option has not been utilized in DataverseNO as we whenever possible prefer to follow more standardized approaches.

According to F3, metadata about Findable data clearly and explicitly include the identifier of the data it describes. F3 is implemented in Dataverse in three ways: the dataset PID is part of the metadata record presented on the dataset landing page; the file PID is part of the metadata record presented on the file landing page; and finally, both dataset and file PIDs are included in exported metadata files. All three implementations are adopted in DataverseNO.

According to F4, Findable (meta)data are registered or indexed in a searchable resource. Datasets published with DOIs in a Dataverse-based repository are harvested and indexed by DataCite Search (n.d.). Through DataCite these metadata are made available to a number of other discovery services, including BASE (Bielefeld Academic Search Engine) (n.d.) and the discovery system used by the libraries at Norwegian universities and university colleges. Schema.org metadata are encoded in Dataverse dataset landing pages and from there indexed by Google Dataset Search (n.d.). In addition to these services, metadata from DataverseNO are also harvested by B2FIND (n.d.), and TROLLing metadata are registered and indexed by the CLARIN Virtual Language Observatory (n.d.).

The current implementation of the Findable principle in the Dataverse application and its adoption in DataverseNO are summarized in Table 2. Green shading indicates (more or less) full implementation or support, whereas orange shading indicates (more or less) lacking implementation or support.

Table 2: The implementation of Findability in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
F1	Support for DOI and Handle	Yes (DOI)
	Always at the dataset level	Yes
	Optionally at file level	Yes
F2	Metadata standards in human- and machine-readable formats: Dublin Core; Documentation Data Initiative (DDI); DataCite; Schema.org	Yes
	Optional custom metadata	No
F3	Dataset PID is part of metadata record presented on Dataset landing page.	Yes
	File PID is part of metadata record presented on File landing page.	Yes
	PIDs are included in exported metadata files.	Yes
F4	DataCite metadata is harvested and indexed by DataCite Search.	Yes. In addition: B2FIND and VLO.
	Schema.org metadata is indexed by Google Dataset Search.	Yes

To be Accessible

The gist of Accessibility is according to Jacobsen et al. (2019, p. 14) the following:

> Protocols for retrieving digital resources should be made explicit, for both humans and machines, including well-defined mechanisms to obtain authorization for access to protected data.

The A part in FAIR consists of two elements, A1 and A2. Being Accessible implies according to A1 that (meta)data are retrievable by their identifier using a standardized communications protocol, i.e. a system of rules that allow information to be transmitted between communication systems. The properties of such protocols are further specified in two sub-principles. First, the protocol is open, free and universally implementable (sub-principle A1.1). Data and metadata stored in Dataverse may be accessed through a number of protocols that are in line with A1.1, including Hypertext Transfer Protocol (HTTP) (2020), rsync (2020) over Secure Shell (SSH) (2020), and Representational state transfer (REST) (2020) via Application programming interface (API) (2020), which provides access through e.g. cURL (2020). All these methods are available for users in DataverseNO. HTTP is the default protocol used when users access (meta)data in DataverseNO, whereas access through the other protocols needs to be clarified with the repository management in advance. According to sub-principle A1.2, Accessible data use a protocol that allows for an authentication and authorization procedure, where necessary. For data access via API, Dataverse supports both session- and API key-based authentication. Both methods are

available in DataverseNO. File access in Dataverse can be restricted, either permanently or during an embargo period. In either case, Dataverse allows depositors to handle authorization by deciding whether potential users should be allowed to request access to restricted files, as well as defining possible terms for access. Also these features are available in DataverseNO, however only for embargoed files. Permanent access restriction is currently not accepted in DataverseNO as the repository by default only accepts data that are intended to be made openly available.

Another important aspect of Accessibility is described in A2, namely metadata being accessible, even when the data for some reason no longer are available. By default, datasets – including the files they contain – cannot be deleted in Dataverse. As previously described, any change applied to a dataset results in a new version of the dataset, leaving all previously published versions still being findable and accessible. There may however occur situations where data have been published that for compelling reasons (e.g. legal issues) should not be openly available. In such cases, the files of a dataset may be deaccessioned, meaning that access to these files is removed. Deaccessioning does not affect the citation metadata of the dataset; the data are thus still findable and citable. After deaccessioning, the metadata include information about why the data are no longer available. The deaccession procedure is available in DataverseNO and has been applied a handful of times.

The current implementation of the Accessible principle in the Dataverse application and its adoption in DataverseNO are summarized in Table 3. As in the previous summary, green shading indicates (more or less) full implementation or support.

Table 3: The implementation of Accessibility in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
A1	Yes	Yes
A1.1	Support for HTTP (W3C), Rsync over ssh (GNU General Public license) RESTful API (e.g., access through cURL)	Yes
A1.2	Authentication API Tokens	Yes
	Authorization service	Yes, but only for embargo. By default, DataverseNO only accepts open data.
A2	A deaccessioned dataset (data not available) is still findable and citable.	Yes
	Metadata includes information about why the data are not available.	Yes

To be interoperable

Interoperability as used in FAIR implies according to Jacobsen et al. (2019, p. 14) that digital resources can be used in the following way:

> When two or more digital resources are related to the same topic or entity, it should be possible for machines to merge the information into a richer, unified view of that entity. Similarly, when a digital entity is capable of being processed by an online service, a machine should be capable of automatically detecting this compliance and facilitating the interaction between the data and that tool.

Interoperable data have three main characteristics, as specified in I1 to I3. First (I1), Interoperable (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. On a general level, this is implemented in Dataverse with Linked Data (2020) support through JSON-LD (2020) for Schema.org, meaning that the general metadata of a dataset and its files are represented in a format that allows the information to be searched for and processed together with other Linked Data supported data. Although available in DataverseNO, this feature would need more systematic adaptation to make published data Interoperable within a Linked Data approach on a larger scale. Currently, Interoperability may work well for general attributes that are provided through metadata automatically encoded in files (e.g. information about file type), but not for more content and domain-specific attributes. Supporting the latter level of Interoperability is a challenge for most general data repositories, given the heterogeneity of long-tail data they host. However, within certain domains, Dataverse offers more elaborate domain-specific Interoperability support. For example, this is true for quantitative social sciences data that may benefit from a DDI (XML) based schema supporting extensive variable metadata. These

possibilities have been applied to some datasets published in DataverseNO.

Another aspect of Interoperability is the use of (meta)data vocabularies that follow the FAIR principles, as stated in I2. FAIR controlled vocabularies and data models may be deployed manually in Dataverse, e.g. as keywords in the general metadata section, usually requiring some guidance from professional data curators. In DataverseNO, there are a few datasets where this possibility has been explored; see e.g. Gammeltoft (2019). FAIR controlled vocabularies may also be implemented in Dataverse through customized metadata schemas, or be specified as prefilled or suggested values in metadata templates (which in Dataverse are called dataset templates). DataverseNO has been making use of the latter approach, particularly in cases where a single project produces multiple datasets with related content, so that vocabulary values to a large extent are common for all datasets (e.g. datasets covering data from time series). By default, however, Dataverse does not support controlled vocabularies and complex ontologies yet. There is ongoing work to implement controlled vocabulary support for several domains in Dataverse, e.g. through the Social Sciences & Humanities Open Cloud (SSHOC) (n.d.). By default, however, Dataverse does not support controlled vocabularies and complex ontologies yet.

The last property of Interoperability – specified in sub-principle I3 – is that Interoperable (meta)data include qualified references to other (meta)data. Such references may be added in two fields of the general metadata schema in Dataverse, one for related data, and another for related materials (other research objects). Currently, information may only be entered as free text into these two fields, and the information is not exported to DataCite. This shortcoming will be remedied in a future version of the Dataverse application.

The current implementation of the Interoperable principle in the Dataverse application and its adoption in DataverseNO are summarized in Table 4. As in the previous summaries, green shading indicates (more or less) full implementation or support, whereas orange shading indicates (more or less) lacking implementation or support. In addition, yellow shading indicates partial implementation or support.

Table 4: The implementation of Interoperability in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
I1	Linked data support with JSON-LD for Schema.org	Partially, for general attributes such as file type
	DDI (XML) as a rich schema to support extensive variable metadata	Partially/in some datasets
I2	FAIR controlled vocabularies and data models may be deployed manually, e.g. in well-curated datasets	Partially/in some datasets
	Custom metadata and metadata template can help.	Partially/in some datasets (metadata template)
	Controlled vocabularies and ontologies not supported by default. But, cf. ongoing work on support for some domains (e.g. SSHOC).	No
I3	DDI schema supports references to other data.	Yes, where applicable
	Not yet supported: structured metadata about related objects included in exported DataCite metadata (coming soon)	No

To be Reusable

Digital resources supporting the last part of FAIR, Reusability, can be described as follows according to Jacobsen et al. (2019, p. 14):

> Digital resources are sufficiently well described for both humans and computers, such that a machine is capable of deciding: if a digital resource should be reused (i.e., is it relevant to the task at-hand?); if a digital resource can be reused, and under what conditions (i.e., do I fulfill the conditions of reuse?); and who to credit if it is reused.

Reusable (meta)data are defined in the FAIR Data Principles as (meta)data that are richly described with a plurality of accurate and relevant attributes. The principle is further elaborated in three sub-principles. Sub-principle R1.1 describes the first characteristic of Reusability as (meta)data being released with a clear and accessible data usage license. Information about data use license or waiver as well as – where applicable – information about data access and terms of use are by default included in the metadata of datasets published in Dataverse. Licenses other than the CC0 1.0 Universal (CC0 1.0) Public Domain Dedication (n.d.) are not predefined, and they are by default not machine-readable. This latter shortcoming has so far not had unbearable consequences for DataverseNO as the repository uses CC0 as its default license, which has been applied to almost all published datasets. Currently, the Dataverse application has no support for explicit and machine-readable licenses for metadata. Therefore, DataverseNO has specified the terms for reuse of metadata on the DataverseNO Metadata Harvesting webpage (n.d.), though only in a human-readable format.

The second characteristic of reusable (meta)data is according to sub-principle R1.2 that they are associated with detailed provenance. Provenance is information about the origin of the data, e.g. how they were obtained and processed, as well as who has been involved in the management of the data. Dataverse has support for rich metadata including information about data authors and other contributors, data providers, data distributors, as well as related data (e.g. used as input data). Any changes made to published datasets are automatically documented through version control in Dataverse. In addition, Dataverse supports the registration of provenance information in a more formal way based on the W3C PROV data model for the interchange of provenance information on the Web (W3C, n.d.). This feature has so far not been made use of in DataverseNO.

The last characteristic of Reusability included in FAIR is described in sub-principle R1.3, which says that reusable (meta)data meet domain-relevant community standards. There is a multitude of (meta)data standards within the different domains of science. Dataverse currently supports a few of them. In addition to the discovery metadata schema, there is quite a substantial schema for social sciences, compliant with parts of the Documentation Data Initiative (DDI). Less rich metadata schemas are available for data from astronomy and astrophysics, and for data from life sciences (Dataverse Metadata References, n.d.). In DataverseNO, these metadata schemas are used in some of the datasets from relevant domains.

There is ongoing work in the Dataverse community to establish metadata schemas including controlled vocabularies for more domains.

In addition to the metadata schemas provided in Dataverse by default, domain-relevant community standards may be implemented by creating custom metadata blocks. As mentioned previously, this option has not been utilized in DataverseNO as we whenever possible prefer to follow standardized approaches.

On file-level, Dataverse automatically extracts metadata from FITS files used in astronomy (FITS (Flexible Image Transport System), n.d.). Currently, DataverseNO does not contain any FITS files.

Also on file-level, Dataverse automatically converts certain tabular file formats (e.g. R frames, Stata, and Excel files) into tab-separated plain text format, which is a more preferred file format for long-term preservation. For these tabular files, Dataverse also extracts upon ingest the variable metadata for each column in the table. This feature has been applied to some of the files in DataverseNO.

In addition to the Reusability implementations that are available by default in the Dataverse repository application, the DataverseNO deposit guidelines include two measures that are essential for increasing the reusability of data. First, data must be submitted in

preferred file formats that are suited for long-term preservation. The guidelines contain a list of preferred file formats for some common file types; for other file types, the file format is assessed during curation. If necessary, file formats are discussed in the DataverseNO curation network and the DataverseNO Advisory Committee before the repository management decide whether the format should be included on the list of preferred file formats.

Second, datasets to be published in DataverseNO must include a ReadMe file containing human-readable information about how to reuse the data, including what in some fields is called a data dictionary (Tierney & Ram, 2020, p. 7). Although ReadMe files are not machine-readable and thus do not support one of the basic intentions of the FAIR Data Principles, they contribute substantially to making research data more reusable by humans. In particular, this is true for data from the long tail of research, which includes fields where domain-specific data and metadata standards still are either non-existent or not established. For this type of long-tail research data, I argue that Reusability for humans has higher priority than Reusability for machines in the implementation phase of the FAIR Data Principles. Paraphrasing Tierney & Ram (2020, p. 1), I also argue that making data reusable falls on a continuum, and entering it should come with feasible barriers.

The current implementation of the Reusability principle in the Dataverse application and its adoption in DataverseNO are summarized in Table 5. As in the previous summaries, green shading indicates (more or less) full implementation or support, yellow shading indicates partial implementation or support, whereas orange shading indicates (more or less) lacking implementation or support.

Table 5: The implementation of Reusability in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
R1		
R1.1	Included in metadata: data use license/waiver; data access and use terms. But, licenses other than CC0 are not predefined and by default not machine-readable.	Yes. Almost all datasets are published under default license CC0.
	By default no support for explicit information about metadata license	Terms for reuse of metadata described on website
R1.2	Rich citation metadata including information about data authors and other contributors, providers, distributors, related data (input data)	Yes
	Versions with changes documented automatically	Yes
	W3C PROV support	No
R1.3	DDI for social science data	Partially/in some datasets
	Metadata blocks for other community standards	Partially/in some datasets
	Ongoing work on support for more domains.	No
	Custom metadata	No
	FITS for astronomy data	N/A (so far)
	File format conversion to reusable formats (tabular)	Partially/in some datasets
		Data in preferred file formats
		Datasets include ReadMe file.

It must be noted that the FAIR Data Principles were not designed as fully operationalized criteria or a checklist, and there currently does not exist a commonly adopted approach or method for how to assess the FAIRness of digital resources or infrastructures. Nevertheless, I argue that the FAIRness evaluation presented in this section may prove useful for the continuous FAIRification efforts in the Dataverse community and beyond.

Table 6 summarizes the results from the FAIRness evaluation of Dataverse and DataverseNO presented in this section. Note that the size of the different elements in the table are not claimed to represent their exact contribution to the overall FAIRness, but they may give a rough indication of the FAIRness of Dataverse and DataverseNO as discussed in this section.

Table 6: Summary of current FAIR implementation or support in Dataverse and DataverseNO

	F				A			I			R		
	1	2	3	4	1.1	1.2	2	1	2	3	1.1	1.2	1.3
Dataverse													
DataverseNO													

= (more or less) full implementation or support
= partial implementation or support
= (more or less) lacking implementation or support
= not applicable

Bearing in mind the mentioned limitations of the FAIRness assessment carried out above, the main conclusions from this section may be summarized as follows: Dataverse and DataverseNO provide strong support for Findability and Accessibility, somewhat weaker support for Reusability and rather weak support for Interoperability. Dataverse and DataverseNO are continuously working to increase their FAIRness support. For instance, DataverseNO has recently been selected as one of 12 repositories to get support from FAIRsFAIR (n.d.) to improve its level of Interoperability.

The FAIRness support provided by applications like Dataverse and by repositories like DataverseNO is fundamentally important for the realization of FAIR. In addition, we need mechanisms which ensure that such support can be provided in a stable way also in the long term. This brings us to the last topic to be discussed in this article, the sustainability and trustworthiness of DataverseNO.

6. How sustainable and trustworthy is DataverseNO?

The realization of FAIR depends to a large extent on an ecosystem of federated infrastructures. In such a system, data are managed in different infrastructures which however all follow common standards to make data discoverable and reusable across infrastructures and scientific domains. Although not explicitly addressed by the FAIR Data Principles, the sustainability of infrastructures enabling FAIR data is recognized as a core issue. In order to be sustainable, infrastructures and services must have appropriate funding. In their report on how to turn FAIR into reality, the European Commission expert group on FAIR data make two priority recommendations on funding. The first one is about providing strategic and coordinated funding (European Commission, 2018, p. 55):

> Funders should adopt a coordinated approach to supporting core infrastructure and services, building on existing investments where appropriate. Funding should be tied to certification schemes, sustainable business models and other community-vetted indicators that demonstrate viability.

The second one is about providing sustainable funding (European Commission, 2018, p. 57):

> Funders who issue requirements on FAIR must provide support to ensure the components of the FAIR ecosystem are maintained at a professional service level with sustainable funding. Service providers should explore multiple business models and diverse income streams.

How does DataverseNO relate to these recommendations? Let us briefly discuss the different elements in turn.

1. Funders should adopt a coordinated approach to supporting core infrastructure and services, building on existing investments where appropriate.

As mentioned earlier, repositories are considered core elements in the realization of FAIR. Given the volume of published datasets, it seems appropriate to say that DataverseNO is a core infrastructure for research data in Norway. Building on existing resources is a key principle of the European Open Science Cloud (EOSC) and should also be applied to national investment strategies. DataverseNO has taken this key principle to heart by – among other things – making use of existing open-source software (Dataverse) and extensive support from the Dataverse developer and user community (to be further discussed below) as well as by drawing on existing human resources and organizational infrastructure at the owner institution and the partner institutions.

2. Funding should be tied to certification schemes, sustainable business models and other community-vetted indicators that demonstrate viability.

a) Funding should be tied to certification schemes

DataverseNO realized shortly after its establishment that certification through a recognized organization is an important and useful way to provide evidence of good quality for depositors, partner institutions and potential users of data published in the repository. After we had worked with our application over a period of two years, DataverseNO earned the CoreTrustSeal in March 2020. The CoreTrustSeal (n.d.) is a quality seal of approval for sustainable and trustworthy research data repositories. Through the CoreTrustSeal certification, data repositories and other infrastructures and services demonstrate that they meet a number of requirements for both technical infrastructure and stewardship model and routines. In total, 15 qualification requirements are included in the CoreTrustSeal certification, and DataverseNO meets 14 of them at the highest level ("The guideline has been fully implemented in the repository"). As part of the certification process, we have developed a comprehensive set of guidelines for the repository, the DataverseNO Policy Framework (n.d.).

b) Funding should be tied to sustainable business models

One of the aspects assessed in CoreTrustSeal certification is the business model which a repository operates on. The business model of DataverseNO builds on common approaches to cooperation between higher education institutions in Norway. The multi-institutional model of DataverseNO has grown out of an institutional service at UiT, realizing the advantage of the network effect as described in Arlitsch & Grant (2018). Instead of establishing similar or even (near-)identical repositories at each and every institution, the repository service first built at one institution was rearranged to meet the demands from a larger community. The repository has been established entirely within existing budget limits at the owner institution and the partner institutions, involving no project-based funding. Instead, funding allocations have been (re-)prioritized, and resources have been gradually scaled in line with increasing demand. Deploying an open-source and free application (Dataverse) to run the main technical infrastructure of the repository renders the costs for technical operation and maintenance at a feasible level. Together with the fee for DOI minting, these costs are shared between the partner institutions. The largest contribution from each partner institution is their investment in institutional curation support for datasets to be published in their institutional collections, as well as their investment in research data management training for support staff and researchers. However, this institutional burden is considerably reduced by the collaborative approach to curational training and knowledge exchange between DataverseNO partner institutions. Furthermore, support from the international Dataverse community is an additional contribution to leverage more powerful research data management support services at the DataverseNO partner institutions.

The Organisation for Economic Co-operation and Development (OECD) has explored the income streams, costs, value propositions, and business models for 48 research data repositories. This survey is summarized in a report from 2017, including a set of recommendations on how to develop sustainable business models for repositories (OECD, 2017a). The authors of the report identify the following broad categories of business models being employed by research data repositories (OECD, 2017a, pp. 38–39):

- Substantially structurally funded (i.e. central funding or contract from a research or infrastructure funder that is in the form of a longer-term, multi-year contract)
- Substantially supported by the host institution (i.e. direct or indirect support from the host institution)
- Substantially depending on data deposit fees (i.e. in the form of annual contracts with depositing institutions or per-deposit fees)

- Substantially funded from access charges (i.e. charging for access to standard data or to value-added services and facilities)
- Substantially supported through contract services or project funding (i.e. charges for contract services to other parties or for research contracts)
- Business models based on a combination of revenue sources

The business model of DataverseNO may be said to be a combination and adaptation of three of these models: 1) host or institutional support; 2) deposit-side contract; and 3) possible future project funding. The OECD report gives an analysis of possible advantages and disadvantages associated with these models. Let us briefly evaluate the DataverseNO business model based on these findings.

Host or institutional support

The main funding source of DataverseNO is institutional support in the form of investment in technical infrastructure at the owner institution, and investment in human resources to provide support at both the owner institution and all partner institutions.

Among the major strengths of host or institutional support models, the OECD report points out the following four (OECD, 2017a, p. 43). The first strength is longer-term sustainability, because universities and other research organizations tend to be long-lived and have robust funding. Both the owner institution and the partner institutions of DataverseNO are state-owned universities and thus part of the national, governmental higher education and research system and under the ultimate responsibility of the Norwegian Ministry of Education and Research. They are all reputable institutions that have existed for many decades – though in some cases not under their current name. Thus, they all are organized and funded in a way that ensures the operation of sustainable services for higher education and research in an enduring perspective.

The second strength of host or institutional support models is convergence of interest between host and repositories. This is typically the case where the data repository aligns with the research strategy of the hosting institution (OECD, 2017a, p. 43). All institutions involved in DataverseNO have recognized Open Science as an important issue in their missions. As a general rule, these institutions retain ownership of the research data produced by their employees, and they thus have a genuine interest in preserving the long-term value of these assets. Therefore, their investment in curational support and participating as partner institutions in DataverseNO is highly convergent with the long-term commitment of the repository.

A third strength of host or institutional models that the report mentions is cost optimisation that may be obtained by “sharing services within the institution” (OECD, 2017a, p. 43). In the case of DataverseNO, cost optimisation is above all a result of the repository

being shared between institutions. This applies above all to costs associated with the development and maintenance of the technical infrastructure, and to a somewhat lesser extent to costs involved in the development of human skills and expertise.

As a fourth major strength of host or institutional models the report considers that "repository support staff can be close to the researchers for on-hand support" (OECD, 2017a, p. 43). The operation – including curation – of institutional collections in DataverseNO is part of the onsite research support services and the institutional management at the DataverseNO partner institutions. This enables repository support staff to provide customized local support for data depositors.

As the major possible weakness of host or institutional models the report reckons that "the focus on the local institutional community […] may lead to fragmentation of domain data, lower levels of curation, and lower interoperability" (OECD, 2017a, p. 43). In my view, this possible weakness is not so much related to the choice of business model, but has rather to do with the question of what type of repository is best suited to meet the different needs of different research communities. DataverseNO is a generic repository and thus contains mostly long-tail data that do not easily fit into trusted domain-specific repositories. However, despite its generic mission, DataverseNO strives to provide domain-specific expertise as far as possible. At the larger partner institutions, data deposited into institutional collections are curated by research data support staff who are subject specialists in addition to being trained in research data management. Special collections of DataverseNO are without exception managed and curated by specialists within the subject in question. As mentioned earlier, the Interoperability support of the repository application is being continuously improved for an increasing number of scientific domains. Overall, DataverseNO has thus taken substantial actions to mitigate the potential risk of fragmentation of domain data, lower levels of curation, and lower interoperability.

Deposit-side contract

In addition to the institutional support as described above the business model of DataverseNO includes a deposit-side contract which is part of the partner agreement. The annual fees from partner institutions cover the shared costs for technical maintenance as well as for training of collections managers, and other support given by the owner institution.

The main strength of business models relying on deposit-side contracts is the potential such repositories might have to achieve economies of scale and cost optimisation (OECD, 2017a, p. 44). Although the institutions currently involved in DataverseNO are not-for-profit organizations, they may profit from effects of economies of scale, and they are of course interested in reducing their costs. As

compared to the establishment of multiple repositories at each of the currently nine partner institutions, the multi-institutional model of DataverseNO obviously contributes to cost optimisation. It is however somewhat uncertain to what extent DataverseNO may be said to have achieved economies of scale.

The OECD report identifies several possible weaknesses of deposit-side contracts (OECD, 2017a, p. 44). Contract-based funding can be unpredictable as commitments are time-limited, e.g. from year to year. DataverseNO partner agreements are terminable. The DataverseNO partner agreement takes into account the consequences of a possible withdrawal of partnership. Datasets from leaving partner institutions may be transferred to other trustworthy repositories on request from the partner institution. In case datasets are to remain in DataverseNO, the leaving institution will have to provide the necessary financial means to ensure the long-term preservation of these datasets. Also, according to the DOI agreement with DataCite, DataverseNO commits to providing access to published datasets for at least ten years after DOI assignment. Given these precautions, the withdrawal of (a substantial part of) partner institutions would of course be unfortunate as it would reduce access to shared resources such as knowledge exchange, as well as reduced opportunities for cost optimisation. It would however not threaten the sustainability of the stewardship of research data still being published in the repository, as data curation still would be the responsibility of the remaining partner institutions. In fact, the basic costs associated with repository maintenance were by and large the same also when UiT operated their own single-institution repository, which was the predecessor of DataverseNO.

Another possible weakness of business models relying on deposit-side contracts is that they may involve relatively high transaction costs in managing contracts (OECD, 2017a, p. 44). DataverseNO keeps administrative costs as low as possible. In the usual case, a partner agreement is signed once, when the partnership is established. After that, partner institutions are invoiced for annual fees. The annual fee is calculated based on a straightforward allocation key. This procedure has proven to be very functional in similar models for collaboration within Norwegian higher education institutions.

The OECD report mentions a third possible weakness of business models based on deposit-side contracts: “In contrast to discipline, subject, or institutional repositories, there may be limited engagement with researchers, as users or depositors” (OECD, 2017a, p. 44). As DataverseNO is a multi-institutional repository, there is no real contrast to institutional repositories with regard to researcher engagement. Research support services at partner institutions are highly involved in the operation of the collections within DataverseNO, particularly through their work with data curation. Partner institutions have well-established venues in place where research data support staff involved in the operation of DataverseNO,

e.g. university library staff discuss data management-related issues with researchers from the different research communities at the institution.

Again, we may note that DataverseNO has taken precautions to mitigate potential risks of the deposit-side contract element of its business model.

Possible future project funding

DataverseNO has been developed and established without any project funding whatsoever. However, we do not rule out the possibility to apply for project-based funding to further develop the repository. A prerequisite for such projects will be that they do not have a negative impact on the long-term sustainability of the repository.

Project funding used in this way will take advantage of the main strength of business models based on this type of funding, namely "support for innovation and development, and opportunities for further developments in the future" (OECD, 2017a, p. 47).

The primary disadvantages of project-based funding identified in the OECD report include "the short-term nature of such funding, lack of flexibility in its spending, and possible diversion of staff and effort from the core tasks and functions of repository operation" (OECD, 2017a, p. 47).

The first two shortcomings are not relevant in the case of DataverseNO, as project-based funding will not be used for operational purposes. The risk of diversion of staff and effort may to a large extent be prevented by hiring temporary staff, a procedure that is commonly practised at Norwegian higher education institutions.

Finally, it should be mentioned that all three funding sources combined in the business model of DataverseNO are compatible with open data principles, as they do not imply any data access charges for users of datasets published in the repository.

Table 7 summarizes the main strengths and weaknesses of the three funding sources combined in the business model of DataverseNO. Green shading indicates strengths, whereas orange shading indicates weaknesses. The rightmost column shows whether the strengths and weaknesses are applicable to DataverseNO. Here, identical shading indicates full applicability, opposite shading indicates lacking applicability, whereas yellow shading indicates partial applicability.

Table 7: Main strengths and weaknesses of the DataverseNO business model

Funding source			Applicable to DataverseNO
Institutional funding	Pros	Compatible with open data principles	Yes
		Longer-term stability	Yes
		Convergence of interest between host and repository	Yes
		Cost optimisation	Yes
		Close to researchers	Yes
	Cons	Fragmentation of domain data	To some extent; but considerable support for domain data
		Lower levels of curation	No
		Lower interoperability	To some extent; but continuous improvement
Deposit-side contract	Pros	Compatible with open data principles	Yes
		Economies of scale	Not relevant
		Cost optimisation	Yes
	Cons	Unpredictable funding	To some extent; but no threat to sustainability
		High administrative costs	No
		Limited engagement with users	No
Project funding	Pros	Support for innovation and development	Yes
		Compatible with open data principles	Yes
	Cons	Short-term nature	No
		Lack of flexibility	No
		Diversion of staff and effort	To some extent; but may be prevented by hiring of temporary staff

This overview shows that the organizational model of DataverseNO has managed to take advantage of all major strengths, and to mitigate or even eliminate the major weaknesses of the funding sources combined in its business model. Notably, by embedding data curation and other support services in the partner institutions, the model combines the scaling advantages of multi-institutional collaboration with the closeness of onsite support for researchers.

Let us now turn to the last part of the strategic and coordinated funding recommendation from the European Commission expert group on FAIR data:

c) Funding should be tied to community-vetted indicators that demonstrate viability

In addition to business models, there is another aspect that should be considered when evaluating the sustainability of research infrastructures, and that is their relation to international networks. The role of such networks is emphasized in another report commissioned by the OECD Global Science Forum (GSF):

> [A]ccess to research data […] is dependent on individual data repositories at the institutional, national and disciplinary levels and on co-ordinated international networks of these repositories (OECD, 2017b, p. 3).

One of the main characteristics of successful networks is that they "define their scope well and avoid the "not-invented-here" or "try-to-solve-it-all" mentality" (OECD, 2017b, p. 23). One such network that has managed to steer clear of both these pitfalls is the Dataverse developer and user community. The most obvious contribution of this network to the success of DataverseNO is of course the Dataverse repository application, which is the technical backbone of DataverseNO. When we started to develop TROLLing back in 2013, Dataverse was virtually the only repository application designed for research data that could be used out-of-the-box. As is the case with any product, the software had (and still has) its shortcomings, but it has always provided solid core functionality for archiving, publishing and citing research data. The last few years, Dataverse has been further developed to meet more and more of the requirements of a FAIR-aligned repository software. As of 11 April 2020, there are at least 55 repositories worldwide that are run on Dataverse (n.d.).

The popularity of Dataverse is reflected in a continuously growing developer and user community. This community consists of hundreds of members, including software developers, researchers, librarians and data scientists. Over 100 people have been and are still contributing to the software code, the main distribution of which is led by Harvard University. Other members of the community contribute with user interface (UI) and user experience (UX) testing and training; presenting and discussing issues in online discussion fora, community calls and the yearly Dataverse Community Meeting; as well as organizing workshops and training sessions (Durand, 2020). Several members of the community are involved in national and international infrastructure projects that aim at developing further different features of the application and/or tools that can be integrated with Dataverse. In 2018, the Global Dataverse Community Consortium (n.d.) was established to provide international organization to existing community efforts and to provide a collaborative venue for institutions to leverage economies of scale in support of Dataverse repositories around the world.

Being embedded in the global Dataverse community has contributed significantly to improve both the technical infrastructure and the operation of DataverseNO, and it has also advanced the further alignment of the repository with international recommendations for sustainable research data stewardship.

To summarize this somewhat longish discussion we may conclude that the way in which DataverseNO is funded and operated is well in line with the first recommendation from the European Commission expert group on FAIR data on strategic and coordinated funding.

The second funding-related recommendation from this expert group concerns sustainable funding. The recommendation contains two parts. Here is the first one (European Commission, 2018, p. 57):

3. Funders who issue requirements on FAIR must provide support to ensure the components of the FAIR ecosystem are maintained at a professional service level with sustainable funding.

This recommendation is being implemented at DataverseNO partner institutions. These institutions require or at least expect their researchers to manage their data in line with the FAIR Data Principles. As consequence, they commit to provide sufficient support to handle these data in a responsible way, e.g. with the services they offer through their institutional collections in DataverseNO. So far, these efforts have been enabled mostly by reprioritizing job tasks and redefining responsibilities at these institutions. As the demand for research data support services increases, institutions will need to consider major reallocations of existing funding to ensure the sustainable operation of these services.

Most of the long-tail research projects carried out in Norway are funded by the host institutions of researchers through ordinary funding. In addition to this ordinary institutional funding, the recommendation above should also be implemented by the Research Council of Norway and other research funding bodies. This can be done in different ways. At the researcher side, funders should cover costs associated with research data management and make it mandatory to include these costs in grant applications for research projects. On the infrastructure side, funders should contribute to the development and operation of not only domain-specific research data infrastructures, but also support services for data from the long tail of research.

The second part of the sustainable funding recommendation from the European Commission expert group on FAIR data is repeated here (European Commission, 2018, p. 57):

4. Service providers should explore multiple business models and diverse income streams.

The business model of DataverseNO already combines different funding sources. As it may become more and more common for institution-external research project grants to cover costs for research data management (see suggestion above), this funding will be an additional income stream for DataverseNO, given that the partner institutions allocate adequate shares of this income to the operation of their institutional collections in DataverseNO.

In the possible case of minor research organisations joining DataverseNO, we also may consider charging data deposit fees from these institutions in exchange for data curation services, at least in an initial phase. However, as a general rule, the organisational model

of DataverseNO requires partner institutions to take responsibility for such core support services for their researchers.

7. Summary and outlook

DataverseNO is a national, generic repository for open research data. This article presents the organization and operation of DataverseNO and gives an evaluation of the repository along three sets of recommendations: the FAIR Data Principles, recommendations on how to turn FAIR into reality, as well as recommendations for sustainable business models for data repositories. DataverseNO provides strong support for Findability and Accessibility, somewhat weaker support for Reusability, and rather weak support for Interoperability. The business model of DataverseNO takes advantage of the strengths and mitigates or eliminates the risks of the funding sources that are combined in the model. As attested by the CoreTrustSeal, DataverseNO has proven to be a sustainable and trustworthy research data repository, primarily for data from the long tail of research.

DataverseNO is continuously working to increase its FAIRness and its sustainability. Other future activities include the improvement of domain-specific support, e.g. by implementing metadata schemas for more domains. As the Dataverse application soon will offer support for sensitive data, we also may want to consider extending the scope of the repository to include this type of data. Finally, DataverseNO will of course remain open for new organizations to join in as partners.

Acknowledgements

Thanks to Helene N. Andreassen, Mercè Crosas, and the editors of this issue of *Ravnetrykk* for useful comments on earlier versions of this article.

Disclaimer

The author is one of the repository managers of DataverseNO.

References

Application programming interface. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Application_programming_interface&oldid=958345761

Arlitsch, K., & Grant, C. (2018). Why So Many Repositories? Examining the Limitations and Possibilities of the Institutional Repositories Landscape. *Journal of Library Administration*, *58*(3), 264–281. https://doi.org/10.1080/01930826.2018.1436778

B2FIND. (n.d.). Retrieved 21 May 2020, from http://b2find.eudat.eu/

BASE (Bielefeld Academic Search Engine). (n.d.). Retrieved 21 May 2020, from https://www.base-search.net/

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world* (pp. XXV, 383). The MIT Press.

CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. (n.d.). Retrieved 23 May 2020, from https://creativecommons.org/publicdomain/zero/1.0/

Christian, T.-M., Gooch, A., Vision, T., & Hull, E. (2020). Journal data policies: Exploring how the understanding of editors and authors corresponds to the policies themselves. *PLOS ONE*, *15*(3), e0230281. https://doi.org/10.1371/journal.pone.0230281

CLARIN Virtual Language Observatory. (n.d.). Retrieved 21 May 2020, from https://vlo.clarin.eu/

Conzett, P. (2019). *Disciplinary Case Study: The Tromsø Repository of Language and Linguistics (TROLLing)*. https://doi.org/10.5281/zenodo.2668775

Conzett, P. (2020). *Research Data Publishing at UiT The Arctic University of Norway* (Version 1) [Dataset]. DataverseNO. https://doi.org/10.18710/JWTJJB

Conzett, P., & Østvand, L. (2018). Støttetenester for forskingsdatahandtering på UiT Noregs arktiske universitet – erfaringar og forslag til beste praksis. *Nordic Journal of Information Literacy in Higher Education*, *10*(1), 65–80. https://doi.org/10.15845/noril.v10i1.283

CoreTrustSeal. (n.d.). Retrieved 21 May 2020, from https://www.coretrustseal.org/

Crosas, M. (2020). Fair Principles and Beyond: Implementation in Dataverse. *Septentrio Conference Series*, *2*, Article 2. https://doi.org/10.7557/5.5334

Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., & Schwartz, A. (2018). *Data policies of highly-ranked social science journals* [Preprint]. SocArXiv. https://doi.org/10.31235/osf.io/9h7ay

CURL. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=CURL&oldid=954043706

Data Documentation Initiative (DDI). (n.d.). Retrieved 23 May 2020, from https://ddialliance.org/

DataCite. (n.d.). [Website]. Retrieved 23 May 2020, from https://schema.datacite.org/

DataCite Search. (n.d.). Retrieved 21 May 2020, from https://search.datacite.org/

Dataverse. (n.d.). Retrieved 21 May 2020, from https://dataverse.org/home

Dataverse Metadata References. (n.d.). Dataverse. Retrieved 23 May 2020, from http://guides.dataverse.org/en/latest/user/appendix.html

DataverseNO Curator Guidelines. (n.d.). Info: DataverseNO. Retrieved 21 May 2020, from https://site.uit.no/dataverseno/admin-en/curatorguide/

DataverseNO Deposit Guidelines. (n.d.). Info: DataverseNO. Retrieved 21 May 2020, from https://site.uit.no/dataverseno/deposit/

DataverseNO Metadata Harvesting. (n.d.). Info: DataverseNO. Retrieved 21 May 2020, from https://site.uit.no/dataverseno/about/#metadata-harvesting

DataverseNO Policy Framework. (n.d.). Info: DataverseNO. Retrieved 21 May 2020, from https://site.uit.no/dataverseno/about/policy-framework/

Dubline Core. (n.d.). Retrieved 23 May 2020, from https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

Durand, G. (2020). Dataverse's Approach to Technical Community Engagement. *Septentrio Conference Series*, *2*, Article 2. https://doi.org/10.7557/5.5424

European Commission. (n.d.). *European Open Science Cloud (EOSC)*. Retrieved 4 April 2020, from https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

European Commission. (2018). *Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data*. Publications Office of the European Union. https://op.europa.eu/s/n1Yo

FAIRsFAIR. (n.d.). Retrieved 21 May 2020, from https://www.fairsfair.eu/

FITS (Flexible Image Transport System). (n.d.). Retrieved 21 May 2020, from https://fits.gsfc.nasa.gov/

Gammeltoft, P. (2019). *The place-name Elverhøy in Norway* (Version 1) [Dataset]. DataverseNO. https://doi.org/10.18710/OG9ARD

GÉANT, & UNINETT. (2019, May). *Why TROLLing is the thing to do for linguists*. In The Field. https://www.inthefieldstories.net/why-trolling-is-the-thing-to-do-for-linguists/

Google Dataset Search. (n.d.). Retrieved 21 May 2020, from https://datasetsearch.research.google.com/

Heidorn, P. B. (2008). Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends*, *57*(2), 280–299. https://doi.org/10.1353/lib.0.0036

Hypertext Transfer Protocol. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Hypertext_Transfer_Protocol&oldid=957536773

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., … Schultes, E. (2019). FAIR Principles: Interpretations and

Implementation Considerations. *Data Intelligence*, *2*(1–2), 10–29. https://doi.org/10.1162/dint_r_00024

JSON-LD. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=JSON-LD&oldid=956136847

Linked data. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Linked_data&oldid=951149328

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396. https://doi.org/10.1037/h0054346

Neylon, C. (2017). Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers. *Research Ideas and Outcomes*, *3*, e14673. https://doi.org/10.3897/rio.3.e14673

Nosek, B. (n.d.). *Shifting Incentives from Getting It Published to Getting it Right*. Retrieved 4 April 2020, from https://osf.io/bxjta/

OECD. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. OECD Publishing. https://doi.org/10.1787/9789264034020-en-fr.

OECD. (2017a). Business models for sustainable research data repositories. *OECD Science, Technology and Industry Policy Papers*, *47*. https://doi.org/10.1787/302b12bb-en

OECD. (2017b). Co-ordination and support of international research data networks. *OECD Science, Technology and Industry Policy Papers*, *51*. https://doi.org/10.1787/e92fa89e-en

re3data.org. (2015). TROLLing; editing status 2020-04-07. *Re3data.Org - Registry of Research Data Repositories*. https://doi.org/10.17616/R3834T

re3data.org. (2017). DataverseNO; editing status 2020-04-07. *Re3data.Org - Registry of Research Data Repositories*. https://doi.org/10.17616/R3TV17

Representational state transfer. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Representational_state_transfer&oldid=956443795

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed., pp. XXI, 551). Free Press.

Rsync. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Rsync&oldid=956572441

Schema.org. (n.d.). Retrieved 23 May 2020, from https://schema.org/

Secure Shell. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Secure_Shell&oldid=957079117

The e-IRG Task Force on the Long Tail of Data. (2016). *Long Tail of Data* (Version 1.74, E-IRG Task Force Document). e-IRG.

http://e-irg.eu/documents/10920/238968/LongTailOfData2016.pdf

The Global Dataverse Community Consortium. (n.d.). Retrieved 21 May 2020, from http://dataversecommunity.global/home

The Social Sciences & Humanities Open Cloud. (n.d.). Retrieved 21 May 2020, from https://www.sshopencloud.eu/

Tierney, N. J., & Ram, K. (2020). A Realistic Guide to Making Data Available Alongside Code to Improve Reproducibility. *ArXiv:2002.11626 [Cs]*. http://arxiv.org/abs/2002.11626

UiT Open Research Data. (n.d.). DataverseNO. Retrieved 21 May 2020, from https://opendata.uit.no/

W3C. (n.d.). *PROV-Overview*. Retrieved 21 May 2020, from https://www.w3.org/TR/prov-overview/

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

The Tromsø Recommendations for Citation of Research Data in Linguistics

Collaboration illustrated

Helene N. Andreassen

Abstract

The Tromsø Recommendations for Citation of Research Data in Linguistics were published in 2019, with a twofold objective: To provide a guide on how to cite research data in linguistics according to good practices, and to contribute to making linguistics a more transparent science. This paper presents the rationale behind the recommendations as well as the development process. The goal is to demonstrate how collaboration, engagement, and different types of expertise are crucial factors to progress towards a culture of sharing of knowledge.

Introduction[1]

The Tromsø Recommendations for Citation of Research Data in Linguistics is a 15-page long document, primarily made up of templates, definitions, and annotated examples. In short: It's a document with a lot of details. Too many details? Not if the objective is to have something judged relevant and useful by all empirical linguists. It should meet the needs of the neurolinguist, the acquisitionist, the variationist, and the language documentarist, to mention but a few linguistic subfields. It should also cover all types of data collected or generated within the field, such as video and audio recordings, transcriptions, glossed text, annotations, experimental data, and introspection.

The short-term goal of the Tromsø Recommendations (Andreassen et al., 2019b) is to provide linguistic researchers, academic publishers, and data repositories with a guide on how to cite research data according to good practices. The long-term goal is to contribute to making linguistics a more transparent science, with a scholarly community adhering to a culture of sharing knowledge.

---

[1] This paper is based on Andreassen et al. (2019a). Thanks to Andrea Berez-Kroeker, Aysa Ekanger and Philipp Conzett for comments on an earlier version of it.

https://doi.org/10.7557/15.5509

This paper tells the story of the Recommendations.

Rationale

Linguistics can be defined as a data-driven social science in which scholars use observations from language use to draw inferences about cognition and social structure. Typical primary data that underpin linguistic analyses are records of language, such as audio recordings, textual productions, judgment data, and eye tracking data, and annotations of these records, such as phonetic transcriptions, frequency counts, acceptability rates, and reaction times. Despite the crucial importance of data in linguistic research, scholars too often fail to cite them properly (Berez-Kroeker et al., 2017; Gawne et al., 2017). This reduces the transparency of the research and thereby also its reproducibility.

Concerns about the use of data in linguistics publications were mentioned already in 1994 by the editor of *Language*, Sarah G. Thomason.

> Because of the traditionally high standards of *Language* regarding linguistic data, I have tried to identify cases where I may need to pay special attention to the accuracy of data: cases where the referees found problems with the data, where the data seems to be incompletely attested, or where a spot check reveals errors. When I began my term as editor, I expected that there would be cases of this kind from time to time. I did not expect that these cases would occur frequently — so frequently, in fact, that the assumption that the data in accepted papers is reliable began to look questionable. (Thomason, 1994, p. 409)

Thomason wrote the editorial note before internet, at a time when there was no infrastructure that could link publications and data, nor any online repositories where data could be archived. As a consequence, publications functioned as the main window to the empirical evidence. Today, technological solutions are in place to allow easy access to research data, but for linguistics, it is not just about improving the transparency and reproducibility of research. It so happens that linguistic data are precious on many levels, also outside the scholarly community. They capture different world views, they capture cultures at given moments as well as their contact over time with each other, and they capture cognitive capacities and variation across language users. This being said, although we now have the tools to create vast amounts of valuable linguistic data, their full potential cannot be explored unless we archive and cite them properly.

Joining forces

The Tromsø Recommendations have their origin in a network that came into existence in 2015 through the project Developing Standards for Data Citation and Attribution for Reproducible

Research in Linguistics (n.d.), funded by the National Science Foundation (NSF). The project invited more than 40 participants from North America and Europe to three multi-day workshops – researchers, archivists, research data management (RDM) specialists, scholarly publishing specialists, institutional management, and funding agencies, with the goal to reveal challenges and possible solutions to data citation and attribution practices. The project culminated in a position statement about reproducibility in linguistics (Berez-Kroeker et al., 2018b), published open access in *Linguistics*[2], and an application to the Research Data Alliance (RDA) to endorse an interest group for linguistic data. The Linguistics Data Interest Group (LDIG) was established in 2017, co-chaired by scholars from three different continents: Helene N. Andreassen from UiT The Arctic University of Norway, PhD in phonology, curator of the Tromsø Repository of Language and Linguistics (TROLLing, n.d.), and responsible for the institutional RDM training programme, Andrea Berez-Kroeker from the University of Hawai'i at Mānoa, associate professor of linguistics and specialist in language documentation and linguistic data management, and Lauren Gawne from La Trobe University, postdoctoral researcher in linguistics with an interest in critical approaches to methodology and RDM.[3]

As stated in the group charter, the overarching objective of LDIG is to "contribute to a positive culture of linguistic data management and transparency in ways that are in keeping with what is happening in the larger digital data management community", with focus on three main topics (taken from Linguistics Data Interest Group, 2017):

- Development and adoption of common principles and guidelines for data citation and attribution by professional organizations, academic publishers, and repositories for language and linguistics.
- Education and outreach efforts to make linguists more aware of the principles of reproducible research and the value of data creation methodology, curation, management, sharing, citation and attribution.
- Greater attribution of linguistic data set preparation within the linguistics profession.

Ever since the beginning, the intention of LDIG has been to function as a scholar-led, community-based project which draws on different members' expertise, experience, and local networks. This way, LDIG hopes to discover the challenges among linguists and meet their needs more efficiently, and also to evoke engagement and a sense of

[2] On 6 December 2018, the position statement was the most downloaded article of the journal (Andrea Berez-Kroeker, p.c.).

[3] Fun fact: In addition to discussions about linguistic data, the LDIG co-chairs learned a lot about collaboration across very different time zones.

commitment among the community members, which today add up to more than hundred people coming from different subfields of linguistics. Another intention has been to work within the frame of RDM specialist communities, to make sure that all outputs from LDIG are up-to-date and in line with good practices. Two examples in this regard are the overarching Research Data Alliance, with its more than 10 000 members, complex thematic community structure, and biannual plenary meetings, and the TROLLing repository, built on the open-source Dataverse platform, with its local operating group consisting of specialists in linguistics, open access, and system development.

Development of the recommendations

It is common practice among teachers to try to evoke engagement among students by teaching them the *why*s before the *how*s. If students understand the purpose of a task, beneficial to themselves and/or to society in general, it may be easier to tackle any challenging, time-consuming, or boring operation required to succeed. Professionals who plan to handle something new are not necessarily very different and may also need internal or external motivation to become engaged in the learning task. If we focus on citation of research data, an obvious external motivation comes from scientific publishers who increasingly require data underpinning research publications be available to the readers.[4] And how do we inform readers about available data? We archive them and cite them in our publication. Internal motivations undoubtedly vary, but everything suggests that more and more scholars become aware of the importance of research transparency and want to do things right.

There are already many documents on the web authored by competent, trendsetting organizations working on open science, available to scholars who want to learn about RDM. One oft-cited example is the FORCE11 Joint Declaration of Data Citation Principles (2014), a set of principles that "cover purpose, function, and attributes of citations". With this landscape as a starting point, the first task of LDIG was to create a document that could speed up the learning process in the linguistic community, an inspirational document on data citation that would speak to all scholars irrespective of their level of RDM skills and competencies. After several asynchronous meetings in the LDIG community, where all members were invited to answer questions and comment on draft versions, the Austin Principles of Data Citation in Linguistics were published in 2018. Based on the content and structure of the FORCE11 Principles, the Austin Principles (Berez-Kroeker et al., 2018a) were formulated with the goal to raise awareness among

[4] In some cases, data may not be shared because of ethical, legal, commercial, or security reasons. In many of these cases, some metadata can still be shared and as such demonstrate the existence of the data.

linguists and encourage them to make informed decisions regarding the accessibility and transparency of their research data. Information about the principles was disseminated rather widely, in local networks as well as on LINGUIST List (n.d.), and people were invited to endorse the principles on a dedicated website. Presently, more than 100 individuals, as well as 10 organizations, have officially stated that they endorse the Austin Principles and that they "support the idea that the data on which linguistic analyses are based are of fundamental importance to the field, and should be treated as such" (Berez-Kroeker et al., 2018a).

In order to help interested scholars work in line with the Austin Principles, the second task of LDIG was to create a document that could serve as a practical guide to citation of linguistic data. One could rightly ask why linguistics would need a separate guide, but as mentioned by the much respected DataCite (n.d.) on their webpage, next to their recommended citation format, different disciplines may come with different challenges. In late 2017, LDIG mounted a working group dedicated to the development of a citation guide for linguistics publications. In addition to the LDIG co-chairs Andreassen and Berez-Kroeker, the group consisted of Philipp Conzett from UiT The Arctic University of Norway, curator of TROLLing with a background in Nordic linguistics, and active in several European RDM projects, and Koenraad De Smedt from the University of Bergen, professor of computational linguistics and coordinator of CLARINO (n.d.).

In a first phase, all LDIG community members were invited to an asynchronous meeting to reflect on metadata and citation practices in linguistics (see Andreassen et al., 2018). Simultaneously, the working group started gathering information about existing citation initiatives within and outside the discipline, in order to identify which citation templates to build on. The group also collected information about metadata and citation practices in repositories for linguistic data indexed in the repository registries re3data (n.d.) and OLAC (n.d.) – a very useful task as it revealed a handful of important challenges for potential reusers of archived data. For instance, the metadata didn't always clearly specify who to cite as authors of the data. Also, in many cases, only one date was entered in the metadata, which because of lack of description of the metadata field, could be interpreted as either the recording date, the deposit date, the publication date, or the last updated date.

In a second phase, three more members joined the LDIG working group, adding more perspectives, competencies, and manpower to the citation project: Lauren Collister from the University of Pittsburgh, director of scholarly publishing with a PhD in sociolinguistics, and specialist in open access and copyright, Christopher Cox from Carleton University, assistant professor of linguistics and much involved in community-based language work, and Bradley McDonnell from the University of Hawai'i at Mānoa, assistant professor of linguistics and specialist in language

documentation. The group continued the work on the citation guide with four audiences in mind: i) academic publishers, who could add or adopt the document into their author guidelines, ii) data repositories, who could check and if needed adjust their metadata templates so as to make archived data properly citable, iii) researchers using data in their work, who could refer to the citation guide in case the author guidelines didn't specify how to cite data, and iv) researchers planning to collect and archive data, who could use the document to determine which metadata to prepare in order to make their data properly citable.

There was quite some discussion in the working group about the level of detail needed for the different audiences, and during the winter 2019, two drafts saw the light, one condensed with only key elements, and one lengthy with more examples and explanations. The group convened (with two members participating via Skype) in Philadelphia in April 2019, on the occasion of the 13th RDA plenary meeting.[5] Not only were people happy to meet and chat in person, but they finally had the chance to sit together and focus. At this point, if some of the readers of the present paper are still unsure whether linguists really need a discipline-specific citation guide: The working group spent three hours – 3 hours – discussing challenges related to the Author and Date fields in the citation template. For each element of the template, no stone was to be left unturned and potentially cause problems for scholars in the future.

Between April and November 2019, the condensed version of the citation guide was sent out for comments twice in the LDIG community, as always with lengthy and fruitful feedback in return. Using the collaborative Google Drive platform, people had access to the same document and could discuss via the Comments function. The citation guide was also sent out for comments to a list of selected linguistic data experts, journal editors, and leaders in the field, who were in a position to encourage adoption or endorsement of the document for their organization or journal after its publication.

In November 2019, a few members of the group convened (one via Skype) in Tromsø on the occasion of the 14th Munin Conference on Scholarly Publishing. All comments from the final feedback round were discussed and incorporated, and after one day of intense work, the first version was ready. Named after the city where they had been finished, the Tromsø Recommendations for Citation of Research Data in Linguistics were shipped off to the Research Data Alliance for endorsement and publication as an official RDA Supporting Output.

As for the lengthy version of the citation guide mentioned previously, this ended up serving as input to Conzett and De Smedt (to appear),

[5] A big thanks to the Linguistic Data Consortium, who generously offered to host the LDIG meeting in their offices.

a chapter on data citation to be published by MIT Press Open as part of a handbook on linguistic data management.

Lessons learned

Ever since the creation of LDIG, core members of the community have continuously worked to raise awareness among linguists, via informal discussions, emails, short presentations at conferences, RDM teaching sessions, training workshops, and summer schools. We do not know the short-term or long-term effects of our efforts, but at least, here are some lessons learned:

- Like researchers in many other disciplines, linguists experience barriers to data citation, such as the lack of awareness, training, standards, and incentives.
- It is important to involve people from different parts of the scholarly community, in order to identify practices and challenges, to get feedback on ongoing work, and eventually to implement good practices in the research and publication process.
- The world is a busy place, which makes it challenging to fully engage people in different sectors. For many researchers, moving from good intentions to practice takes time. For many academic publishers, other aspects of the publishing process are considered more pressing. For many repositories, good practices for data citation are not contained in the metadata and documentation guidelines.
- Continuous outreach seems to move things (slowly) forward, but concrete outputs, such as the Austin Principles and the Tromsø Recommendations, are key. Also, outreach must happen in the right context, with enough time for presentation and Q&A. Finally, getting the right people on board, decision-makers or trend-setters in the community, is very useful for planning ahead.

Next step

Today, the world has focus on dealing with the coronavirus disease and rapid sharing of knowledge across research institutions world-wide is essential. When the world at some point returns to a more normal state, we may hope that junior as well as senior researchers have gained increased awareness of the importance of research transparency and openness. The Tromsø Recommendations will perhaps not directly contribute to save lives, but they may nevertheless function as an important tool for linguists who wish to carry out transparent and open research.

LDIG therefore now enters a new phase, where education of young researchers and outreach to different sectors of the linguistic community are put in the center of attention. We will simultaneously continue to work on specific topics within RDM, such as metadata and archiving, in order to ensure that the Austin Principles, the

Tromsø Recommendations, and our tips and advice in general are continuously up-to-date and in line with good practices.

Concluding remarks

This paper has told the story of the Tromsø Recommendations for Citation of Research Data in Linguistics, a product that has strongly benefited from the engagement, experiences, skills and competencies in the LDIG community and its associated networks. I hope this paper may encourage practitioners in other fields to initiate similar advancements, if possible within the frame of RDA. I also hope it may inspire decision-makers and publishers to actively collaborate with and support scholar-led initiatives working toward better research practices.

Acknowledgements

Many thanks to the participants in the Data Citation and Attribution project, the Data Science for All of Linguistics project, members of the RDA LDIG, and attendees at our previous workshops, courses and presentations for fruitful discussion.

This material is based upon work supported by the National Science Foundation under Grants No. 1447886 and 1745349. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Andreassen, H. N., Conzett, P., De Smedt, K., Berez-Kroeker, A. & Gawne, L. (2018). *Data citation and metadata standards in linguistics.* Paper presented at the LDIG working session during the RDA 11th Plenary Meeting, 21–23 March 2018, Berlin, Germany. Retrieved from https://hdl.handle.net/10037/16556

Andreassen, H. N., Berez-Kroeker, A., Collister, L., Conzett, P., Cox, C., De Smedt, K., Gawne, L. & McDonnell, B. (2019a). *Data citation in linguistics publications: A scholar-led, community-based initiative.* Paper presented at the 14th Munin Conference on Scholarly Publication, 27–28 November 2019, Tromsø, Norway. https://doi.org/10.7557/5.4876

Andreassen, H. N., Berez-Kroeker, A. L., Collister, L., Conzett, P., Cox, C., De Smedt, K., McDonnell, B. & Research Data Alliance Linguistics Data Interest Group. (2019b). *Tromsø Recommendations for Citation of Research Data in Linguistics*. https://doi.org/10.15497/rda00040

Berez-Kroeker, A. L., Gawne, L., Kelly, B. F. & Heston, T. (2017). *A survey of current reproducibility practices in linguistics journals, 2003–2012*. Retrieved from https://sites.google.com/a/hawaii.edu/data-citation/survey

Berez-Kroeker, A. L., Andreassen, H. N., Gawne, L., Holton, G. Kung, S. S., Pulsifer, P. Collister, L. B., The Data Citation and

Attribution in Linguistics Group & the Linguistics Data Interest Group. (2018a). *The Austin Principles of Data Citation in Linguistics, version 1.0.* Retrieved from http://site.uit.no/linguisticsdatacitation/austinprinciples

Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., Pulsifer, P., Beaver, D. I., Chelliah, S., Dubinsky, S., Meier, R. P., Thieberger, N., Rice, K. & Woodbury, A. C. (2018b). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics 56(1)*, 1–18. https://doi.org/10.1515/ling-2017-0032

CLARINO: Common Language Resources and Technology Infrastructure Norway. (n.d.). Retrieved 18.05.2020 from https://clarin.w.uib.no/

Conzett, P. & De Smedt, K. (to appear). Guidance for citing research data. In A. Berez-Kroeker, B. McDonnell, E. Coller & L. Collister (Eds.), *The Open Handbook of Linguistic Data Management.* MIT Press Open.

Data Citation Synthesis Group, Martone, M. (Ed.) (2014). *Joint Declaration of Data Citation*. San Diego, CA: FORCE11. https://doi.org/10.25490/a97f-egyk

DataCite. (n.d.). *DataCite – Cite your data.* Retrieved 27.04.2020 from https://datacite.org/cite-your-data.html

Developing Standards for Data Citation and Attribution for Reproducible Research in Linguistics. (n.d.). Retrieved 27.04.2020 from https://sites.google.com/a/hawaii.edu/data-citation/

Gawne, L., Kelly, B. F., Berez-Kroeker, A. L. & Heston, T (2017). Putting practice into words: The state of data and methods transparency in grammatical descriptions. *Language Documentation & Conservation 11*, 157–189. Retrieved from http://hdl.handle.net/10125/24731

Linguistics Data Interest Group. (2017). *Linguistics Data Interest Group charter statement*. Retrieved from https://www.rd-alliance.org/groups/linguistics-data-ig

LINGUIST List. (n.d.). Retrieved 19.05.2020 from https://linguistlist.org/

OLAC: Open Language Archive Community. (n.d.). Retrieved 27.04.2020 from http://www.language-archives.org/

Re3data: Registry of Research Data Repositories. (n.d.). Retrieved 27.04.2020 from https://www.re3data.org/

Thomason, S. G. (1994). The Editor's Department. *Language 70(2)*, 409–413. Retrieved from https://www.jstor.org/stable/415877

TROLLing: The Tromsø Repository of Language and Linguistics. (n.d.). Retrieved 18.05.2020 from https://trolling.uit.no/

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Open Polar

A mid-way report toward a unique service

Tamer Abu-Alam

Abstract

Data from the Polar Regions are of critical importance to modern research and decision makers. Regardless of their disciplinary and institutional affiliations, researchers rely heavily on the comparison of existing data with new data sets to assess changes that are taking effect. However, in a recent survey of 113 major polar data providers, we found that an estimated 60% of the existing polar research data is unfindable through common search engines and can only be accessed through institutional webpages. This raises an awareness sign of the need of the scientific community to harvest different metadata related to the Polar Regions and collect it in a homogenous, seamless database and making this database available to researchers, students and publics through one search platform.

This contribution describes the progress in an ongoing project, Open Polar, started in 2019 at UiT The Arctic University of Norway. The project aims to collect metadata about all the open-access research data, articles and other scholarly documents related to the Polar Regions in a homogenous and seamless database. During the first six months of the project, the beta version of the user-interface was established, with a search by map and an advanced search function. An extensive geo-database that includes thousands of polar locations and their geographic information was collected from different sources. The geo-database together with a list of keywords (i.e. on sources, indigenous peoples, languages and other polar-related keywords) will be used in the filtration process.

A Reference Board was formed, and the first board meeting took place in April 2020. The geographic definition of "Polar Regions" was defined in order to include most of the current geographic definitions of "Arctic". The project is still facing some challenges that include for example integration with non-standard data sources who do not use Dublin Core Metadata schema, or are not harvestable through the Open Access Initiative's standard protocol for harvesting (OAI-PMH).

https://doi.org/10.7557/15.5505

Introduction

The Polar Regions (i.e. the Arctic and Antarctic) are places where global processes interact across the Earth and therefore these regions are considered as barometers to measure the health of the Earth, and as key areas to study the global changes in the Earth's ecosystem (e.g. IPCC, 2007, 2013; Beck et al., 2014). Several recent climate impact assessments on the ecosystem of the Polar Regions (e.g. AMAP, 2011, Forbes, 2011, Turner et al., 2009, 2013, Arctic Council, 2013, Krupnik et al., 2011) draw a consistent pattern of climate-driven environmental, societal, and economic changes in recent decades. These assessments rely heavily on the comparison of existing data with new data sets to assess changes that are taking effect.

Despite its importance, there are massive access barriers to the research data. Tenopir et al. (2011) concluded that 871 (i.e. 67%) out of 1300 researchers pointed to a lack of access to data generated by other researches or institutes in Europe. A similar conclusion was drawn on the polar-related research data, where a findability gap of about 60% was found (Abu-Alam, 2019). Moreover, a study by Johnson et al. (2019) on the Arctic region showed that social sciences and indigenous studies have a findability gap around 84%. Thus, this findability gap is not connected only to the Polar Regions, but similar gaps can be observed on different research disciplines covering different geographic locations. However, the present contribution focuses on the findability gap of polar-related research data.

This contribution describes an ongoing project, Open Polar (https://site.uit.no/open-polar/) started in 2019 at UiT The Arctic University of Norway. The project aims to collect metadata about all the open-access research data and documents related to the Polar Regions in a homogenous and seamless database and making this global output of open-access material available through an inter-active user interface. The project will help to make research data and documents considered to be of relevance to the polar regions more visible and searchable to the end-users and as a result reducing the findability gap.

Open Access research data – a historical view

"Open Access" is a term that has come to mean a set of policies about, as well as methods for publishing, archiving, and disseminating scholarly works. The history of Open Access can be traced back to the end of the 20th century, when university librarians around the world faced a problem known as the "serials crisis". The serials crisis occurred when the subscription costs for publications rose much faster than inflation. In order to balance the limited funds available to the libraries with these subscription high costs, the libraries were forced to make difficult choices between different journals. The serials crisis was the engine behind a larger movement to make scientific research more easily available, and at a lower price.

Initially, the term was related only to free access to peer-reviewed literature (e.g. Budapest Open Access Initiative, 2002). In 2003, through the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities", the definition was considered to have a wider scope that includes raw research data, metadata, source materials and scholarly multimedia material.

In line with the goals expressed in the Berlin Declaration, UiT The Arctic University of Norway took the initiative to make open access research documents about the Arctic more visible and searchable through a search service known as "High North Research Documents", launched in 2012 (Longva and Høydalsvik, 2012; Longva, 2011). Some other services dealing with polar research data were established around this time (e.g. Polar Data Catalogue: https://www.polardata.ca/), however the High North Research Documents stands up comparing to other services in terms of the contents, the number of records and its easy to use user-interface.

In 2016 and with the advanced technology in digitalization and cloud storage, the scientific society moved forward in their revolution supporting Open Access. Wilkinson et al. 2016 introduced the "FAIR Data Principles" as a set of guiding principles to make data findable, accessible, interoperable and reusable. Although many research institutes and researchers start to move toward this new digital era by uploading their research data to different institutes' repositories, most of these research data are not findable through standard search engines (e.g. Google Scholar). Consequently, a considerable amount of research data uploaded to institutes' repositories is consistent with the second principle of FAIR (i.e. accessible) but not with the first principle (findable) nor with the third or the fourth principles (interoperable, reusable). Therefore, new services that aim at making research data and documents more visible and searchable are needed.

The need for a new service in the polar sciences

Although the High North Research Documents is considered as an advanced step toward making Arctic-related records more searchable and findable through an easy to use user-interface that allows searching an extensive database that contains more than 1,200,000 open access records, a new service is needed in order to widen the scope to cover all the Polar Regions and to cover both research data and research articles.

A pilot study (Open Arctic Research Index) at the UiT conducted a survey in order to measure the needs of the scientific community for a new service that will provide the community with access to the records published on the Polar Regions through one search platform. The main outcomes (Abu-Alam 2019), as well as the steps that are currently being made to remedy the situation, are described below.

60% findability gap of the polar records

In the survey, a number (i.e. 113) of major and trusted institutes / organizations / research units (hereafter described as "metadata providers" or simply "providers") were mapped. The survey focused only on providers who are dealing exclusively or primarily with polar sciences. The different metadata providers were classified based on their findability through the Bielefeld Academic Search Engine (BASE) and High North Research Documents into three categories:

1- Metadata providers not included in BASE and High North Research Documents (58 providers)
2- Metadata providers included in BASE and High North Research Documents (21 providers)
3- Metadata providers included (partially, as publisher) in BASE and High North Research Documents. These partially included providers do not give full access to all their catalogs (34 providers)

The High North Research Documents and the database of BASE were used here as a basis of our analysis and as examples of common search engines, however, major search engines (e.g. Google) were used to validate the results. The reasons for choosing the High North Research Documents and BASE as examples of the common search engines are that these search engines contain more than 1,200,000 records from the Polar Regions and therefore, they represent highly trusted infrastructures.

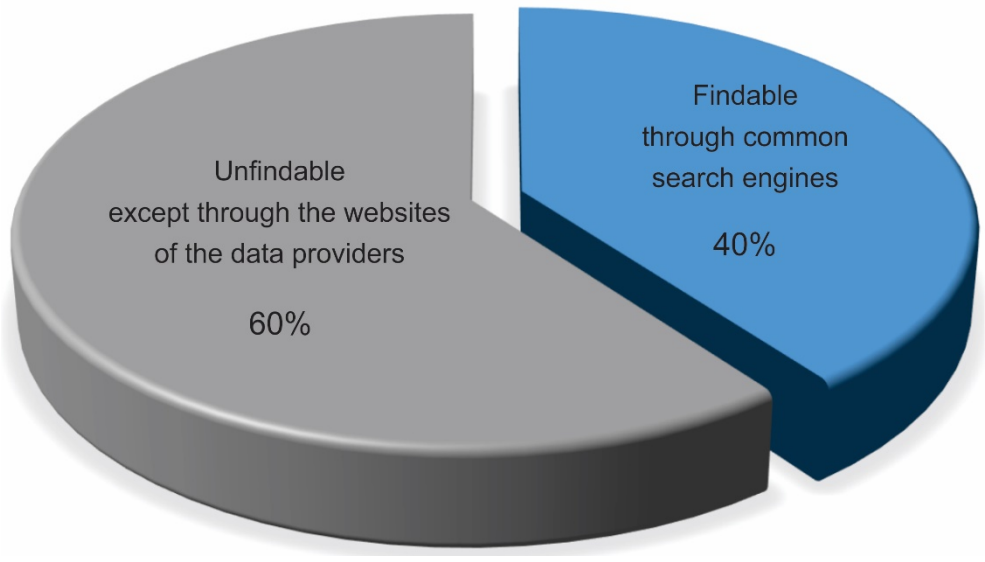


Figure 1. There is a 60% findability gap of the polar-related records based on the records numbers of 113 metadata providers.

The number of records from metadata providers that are not included and partly included in BASE and High North Research Documents are used to indicate the findability gap of the polar records. Interestingly, the numbers of records that are unfindable through common search engines represent around 60% of the total numbers

of the polar records (Fig. 1). This high percentage raises a series of questions. One of these questions is "why is such a high percentage unfindable although the survey targeted major and trusted data providers?" However, discussing such questions is beyond the current contribution. Instead, we will focus here on how to develop a service with the purpose of rendering polar research more visible and retrievable to the scientific community as well as to the interested public, teachers, students and public services.

Arctic versus Polar: broadening the scope

It is worth mentioning here that the sample providers used in the survey are not focused only on research activities in the Arctic area, but the sample covered providers active in both Polar Regions (i.e. Arctic and Antarctic). Therefore, the survey raised a question, namely: should the new service focus only on Arctic Open Access records? Or, should the new service cover all the Polar Regions?

The Arctic and the Antarctic regions are analogue to each other. Many researchers who are working in the Antarctic region are active in the Arctic region as well. Different research funding agents (including European/Scandinavian states funding agents) maintain and operate a large number of observatories and research stations both in the Arctic and Antarctic and as a result the research activities in both regions are linked and relevant (e.g. Ricci & Egerton, 2010; Norsk Polarinstitutt, strategi 2019–2024). In order to provide the scientific communities with a service that covers their needs, the service has to cover all the Polar Regions.

Open Polar: a unique service

The project aims to maximize the number of findable records, therefore metadata from different providers need to be added. The team of Open Polar targets metadata from different national and international metadata providers. By adding the records of these metadata providers, the numbers of the records in the Open Polar's database will be extended to the maximum (as possible), covering a wide range of subjects.

Most of the metadata providers host records from the Polar Regions but also from non-Polar Regions. As a result, a filtration algorithm is needed in order to keep only the relevant records. The filtration algorithm will be conducted by searching the harvested records for one or more keywords. A list of Polar Regions-related keywords will be prepared for this purpose. The relevant records will be saved in a database (hereafter: "Open Polar's database").

Open Polar's database will be structured according to the Dublin Core Metadata Schema. The Dublin Core Schema is a small and simple set of vocabulary terms that can describe any record (i.e. digital or physical record). The schema includes 15 metadata terms, known as the Dublin Core Metadata Element Set. Although the Dublin Core

Metadata is a simple schema, combination of different metadata vocabularies can be used to describe highly complicated records.

The records saved in Open Polar's database will be presented to the end-users using an interactive open-source user-interface. The Open Polar project team decided to use VuFind as a basis of the user-interface. The VuFind is an open-source library resource portal developed by Villanova University's Falvey Memorial Library and maintained by a large community that keeps the source code up to date. The VuFind characterizes by several options that will produce an attractive user-interface. These options include, for example:

- a map search; an option that allows the users to search Open Polar's database using the geographic location of the records;
- search with Faceted Results where the search system allows the user to search from a basic search box and then to be able to narrow down the results by clicking on the various facets of the results;
- advanced search where the users can focus their search to a set of search words in specified metadata fields and/or exclude some records by excluding un-wanted search words;
- save search and organized lists where the users have the ability to save their research results to their own customizable lists. The lists can be retrieved at any time. This helps to eliminate the need for desktop-based citation management software that tends to be too difficult for basic users;
- export the citation of records; the VuFind is able to export the citation of the searched records into several citation format including the APA, Chicago Style and/or the MLA. These exported citations can be saved in a proper format that suite different desktop-based citation management software (e.g. Endnote, RefWorks);
- author biographies where the users can learn more about the authors with contextual information and see all of the records that they have produced;
- URLs to the source: the VuFind is able to present one or more URL to the source of the records which allows the users to retrieve and access the original articles or data (this is an important function since the Open Polar aims to harvest only metadata and this function will allow the users to access the data itself);
- access original data: although we target to add an URL to each record (see the previous point), the VuFind has many APIs to interact with the search, the data and many other features. Users can syndicate the record data with other search services via an OAI server. Users can search using VuFind's algorithms via OpenSearch. And if users want complete access to the indexed data, they can interact with Solr, VuFind's backend search and index engine.

- Internationalization, in the sense that the interface of the VuFind has translations available in several languages including Norwegian, Sámi, Brazilian, Portuguese, Chinese, Dutch, English, French, German, Japanese, Spanish, Arabic and more.
- autocomplete/autosuggestion function is an algorithm that allows the end-users to perform search even if they misspelt the search word.

Location name	Control_name1	Latitude	Longitude
Amundsen Basin	antar*	-60	-115
Attu Station	Alaska	52.84642	173.1861
Attu Mountain	Alaska	52.89139	172.9322
Attu Island	Alaska	52.9025	172.9094
Akimiski island		53.0084	-81.2747
Atlin Lake	British Columbia	59.53194	-133.722
Atlin Lake	Yukon	60.01333	-133.831
Attu	Finland	60.18333	22.33333
Alaska range		62.58333	-153.083
Alpha ridge		63.52611	-151.017
Cape Dorset		64.17851	-76.482
Attu	Greenland	67.93552	-53.5791
Cambridge Bay		69.03665	-105.13
Andfjorden		69.16667	16.33333
Brooks Range	Alaska	69.20222	-143.802
Balsfjord		69.24024	19.22653
Båtsfjord		70.63506	29.72149
Babbage Bay		71.56798	-93.5837
Baffin Island		72.23384	-88.8245
Bylot Island		73.25996	-78.8049

Table 1. An example of the geo-database showing the location names, control_name1 and the latitudes and the longitudes.

In order to allow the end-users to use map search, most of the records in Open Polar's database should have geographic information (i.e. latitude and longitude). However, most of the records are harvested without having their geographic information. As a result, there is a need to create a new algorithm that allows an automatic recognition of the geographic information. In order to do that a database (hereafter: "geo-database"; Table 1) will be established, containing all the existing polar locations with their geographic information. The

algorithm works by searching each record for a polar location (e.g. name of a city, a mountain or a geographic feature). Once the polar location is found, the algorithm will set the pre-defined geographic information of this location to the record's metadata.

A unique service: Reference Board

To ensure that the content of the service is up-to-date and meets the requirements of the scientific community, a reference board consisting of a selected group of researchers and scientists from UiT The Arctic University of Norway and the Norwegian Polar Institute (NPI) has been appointed. The reference board gives advice on strategic matters as well as content selection and functionality aspects of the search engine as such. Since the proposed service is multi-disciplinary, the reference board covers social sciences, humanities, physical and life sciences. The input from the reference board has already proven its worth and helps make Open Polar a unique service in comparison with similar common search engines that do not rely on expert researchers.

A unique service: National-International Board

A board representing different national and international partners and different polar nations is suggested to replace the reference board by the end of the project period and before launching the service. The board can highlight metadata providers from different countries that are not covered by the service and help pave the way for the inclusion of precious records that are currently unfindable. Moreover, the members of the national-international board will help with the dissemination of the service in their respective networks.

Additional subservices

The existing service at the UiT (i.e. High North Research Documents) is used by several universities for educational purposes. However, the way of presenting the records in this service is streamlined mainly for research and not for educational purposes. To establish an attractive and comprehensive service, an educational platform is planned to be added to the Open Polar as a subservice. The needs of this educational platform have recently emerged worldwide due to the pandemic of the Coronavirus (i.e. teaching from home). In this context, we suggest that this subservice can be performed in two steps:

1) collect a series of online lectures and talks on the polar sciences and make it available and searchable to students. The Arctic Massive Open Online Course (Arctic MOOC), which is currently under construction at UiT, is a candidate to be included in this step;
2) simplify some selected research data to be used for educational purposes.

Other subservices are suggested to be included in the Open Polar. These subservices include (Fig. 2): 1) archiving original data from the Polar Regions; and 2) creating a research platform. Although these subservices, including the education platform, will enrich the final service of the Open Polar, the project team is focusing on the main part of the service (i.e. harvesting Polar Regions-related records and making them more visible and searchable). Therefore, discussing these subservices is beyond the current contribution, but more information about these subservices are found in Abu-Alam (2019).

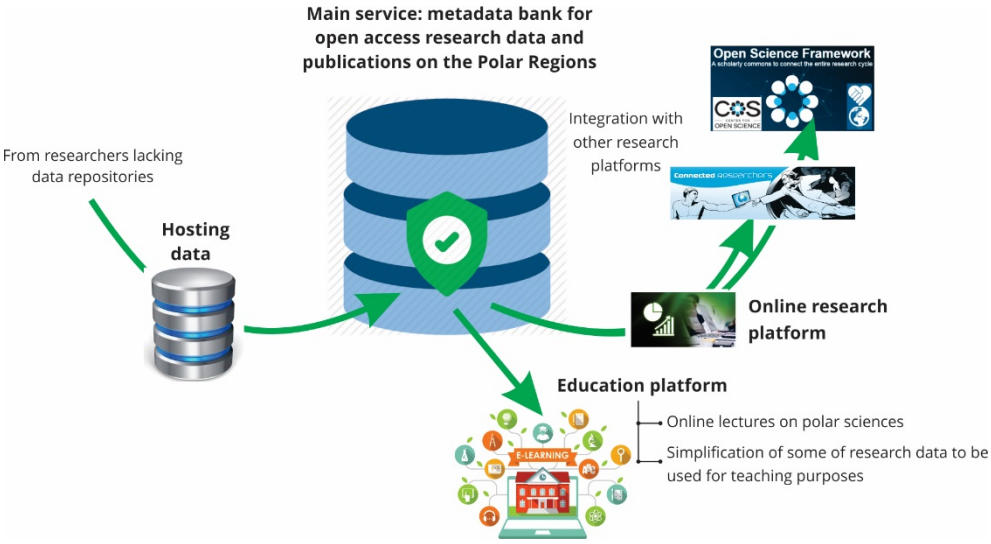


Figure 2. The integration between the main service (i.e. make open-access polar records more visible and findable) and the three additional subservices (modified after Abu-Alam, 2019).

Progress, challenges and planned steps: a conclusion

During the first six months of the project progress has been achieved. The beta version of the user-interface was established and put online (https://james.ub.uit.no/vufind/; Fig 3). In this beta version, the search by map and the advanced search tools are included. An extensive geo-database that includes thousands of polar locations and their geographic information was created (e.g. Table 1). Although the geo-database needs to be cleaned up (e.g. remove unrelated locations and duplications and add some locations with different languages), a simple version of the geo-database was tested successfully. The geo-location database will be used also in the filtration process.

In March 2020, the Reference Board was formed, and the first board meetings took place in April 2020. One of the challenges that face the Open Polar is the geographic definition of the word Polar. The Reference Board decided that the geographic definition of the "Polar Regions" should include most of the current geographic definitions of "Arctic". In addition, the board emphasized that the Open Polar project must consider different national definitions of the "Polar

Regions", including the Norwegian term "nordområdene" (high north), which includes subarctic as well as arctic areas.

Although the Dublin Core Metadata is a standard schema used in the communication among different metadata providers, many providers do not use this schema. Even some of the providers who are using the schema, they do not use it in a proper way (i.e. some information is entered in wrong fields). It is a challenge to use metadata not organized in a proper way. Therefore, a combination of MySQL and Solr technology together with GO and php programming codes will be used in order to read, organize and re-sort the metadata fields.

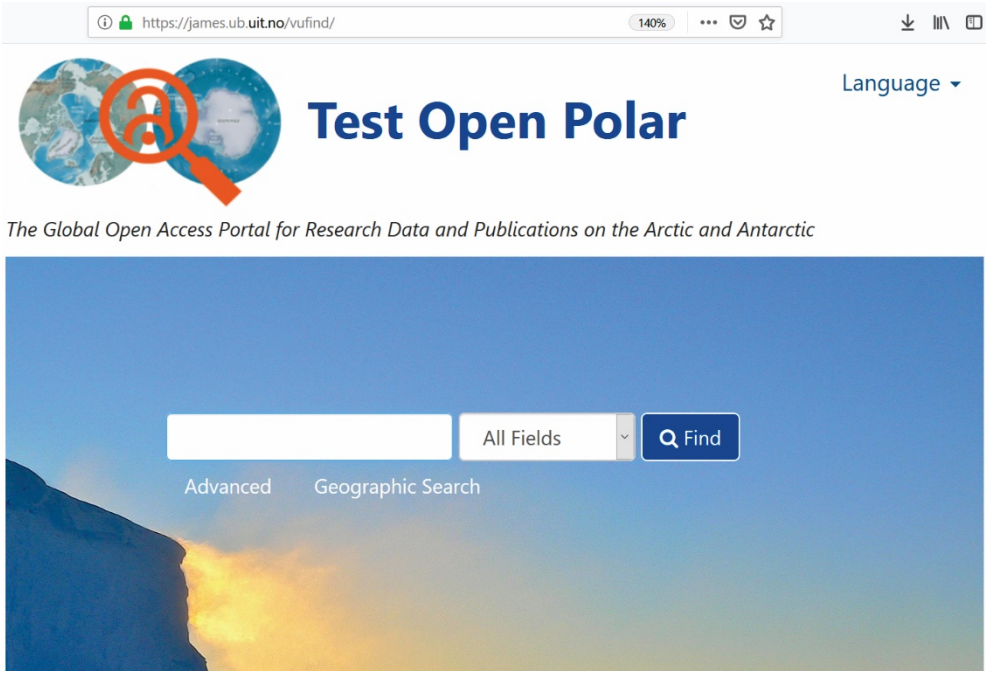


Figure 3. Screenshot shows the beta version of the Open Polar search tool.

It is our plan that by September 2020, all the databases that will be used in the filtration process and identification of geographic information of each record will be completed. In addition, most of the errors and defects in the beta version will be fixed. Before the end of 2020, the filtration algorithm will be tested, and the results will be uploaded to Open Polar's database after adding the missing geographic information. Integration with non-standard metadata sources will be our target for November and December 2020.

Although the Coronavirus pandemic affects our outreach plan for 2020, an extensive promotion plan is planned for the first half of 2021. The promotion plan will be associated with attracting new metadata providers to share their metadata for inclusion in Open Polar. The preliminary plan is to launch the search engine by the second half of 2021.

Open Polar Team

On a daily basis, the Open Polar project depends on a working team. In addition, the project gets occasional advice from the reference

board, whereas major strategic decisions are made by a steering group. As coordinator of the Open Polar project, I would like to thank my colleagues at the Open Polar.

Working team:

- Stein Høydalsvik – the founding father and architect behind both the pilot project OpenARI and the ongoing Open Polar project, UiT university library.
- Leif Longva – central to the High North Research Documents and an important counsellor to Open Polar, UiT university library.
- Karl Magnus Nilsen – taking care of several aspects of the technical parts of the service including filtration algorithm and setting up the user-interface, UiT university library.
- Obiajulu Odu – deals with the technical parts of the service including harvesting metadata using the application programming interface (API), UiT university library.

Steering group:

- Johanne Raade (library director, UiT)
- Per Pippin Aspaas (head of research and publishing support, UiT university library)
- Stein Tronstad (Section leader, NPI)

Reference board:

- Jørgen Berge (UiT)
- Else Grete Broderstad (UiT)
- Alfred Hanssen (UiT)
- Hans-Kristian Hernes (UiT)
- Henry Patton (UiT)
- Lena Seuthe (UiT)
- Øyvind Lundesgaard (NPI)
- Ylva Ericson (NPI)

Bibliography

Abu-Alam T.S. (2019). Open Arctic Research Index: Final report and recommendations. *Septentrio Reports (No. 3).* https://doi.org/10.7557/7.4682

AMAP (2011). AMAP Assessment 2011: Mercury in the Arctic. *Arctic Monitoring and Assessment Programme (AMAP), Oslo, Norway. xiv + 193 pp.* ISBN 978-82-7971-068-4

Arctic Council (2013). Summary for policy-makers. Arctic Resilience Interim Report. *Stockholm Environment Institute and the Stockholm Resilience Centre to the Arctic Council's 8th Ministerial Meeting in Kiruna, Sweden.* http://hdl.handle.net/11374/1629

Beck, I., Huffman, L. T., Xavier, J. C. C. & Walton, D. W. H. (2014). Education and Polar Research: Bringing Polar Science into the Classroom. *Journal of Geological Resource and Engineering 4, 217-221.* https://doi:10.17265/2328-2193/2014.04.004

Forbes, D.L. (2011). State of the Arctic Coast 2010 – Scientific Review and Outlook. *International Arctic Science Committee, Land-Ocean Interactions in the Coastal Zone, Arctic Monitoring and Assessment Programme, International Permafrost Association. Helmholtz-Zentrum, Geesthacht, Germany, 178 p.* http://arcticcoasts.org, ISBN 978-3-9813637-2-2

IPCC (2007). Climate Change 2007: Synthesis report, edited by Pachauri, R. K. & Reisinger, A. *IPCC, Geneva, Switzerland, 104 pp.* https://www.ipcc.ch/report/ar4/syr/

IPCC (2013). "Summary for Policymakers". In Climate Change 2013: The Physical Science Basis, edited by Stocker, T.F., & Qin. D. H. *Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 1535 pp.* https://www.ipcc.ch/report/ar5/wg1/

Johnson, N., Druckenmiller, M. & Pulsifer, P. (2019) The Exchange for Local Observations and Knowledge of the Arctic (ELOKA): Working towards interoperability of community-based data platforms. *Polar Data Forum*

Krupnik, I., Allison, I., Bell, R., Cutler, P., Hik, D., López-Martínez, J., Rachold, V., Sarukhanian, E. & Summerhayes, C. (2011). Understanding Earth's Polar Challenges: International Polar Year 2007–2008. *University of the Arctic, Rovaniemi, Finland / CCI Press, Edmonton, Alberta, Canada and ICSU/WMO Joint Committee for International Polar Year 2007–2008, 724 pp.* ISBN 978-1-896445-55-7

Longva, L. & Høydalsvik, S. (2012). High North Research Documents: your source for research documents on the North. *Polar libraries bulletin 68, 7-9.* http://hdl.handle.net/10037/4733

Longva, L. (2011). NAROS: Northern Areas Open Scholarly Documents. http://DOI:10.7557/13.1312

Norsk Polarinstitutt, strategi 2019 – 2024. *Retrieved from* http://www.npolar.no/npcms/export/sites/np/files/vedlegg/strategi.pdf

Ricci, C.-A., Egerton, P. (2010). European Research in the Polar Regions: Relevance, strategic context and setting future directions in the European Research Area. *Edited by the ESF European Polar Board for European Science Foundation, 24 pp.* ISBN: 978-2-918428-30-5

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011) Data Sharing by Scientists: Practices and Perceptions. *Practices and Perceptions. PLOS ONE 6(6): e21101.* https://doi.org/10.1371/journal.pone.0021101

Turner, J., Bindschadler, R.A., Convey, P., di Prisco, G., Fahrbach, E., Gutt, J., Hodgson, D.A., Mayewski, P.A. & Summerhayes, C.P. (2009). Antarctic climate change and the environment. *Cambridge: Scientific Committee on Antarctic Research, 526 pp.* ISBN 978-0-948277-22-1

Turner, J., Bracegirdle, T.J., Phillips, T., Marshall G.J. & Hosking, J.S. (2013). An initial assessment of Antarctic sea ice extent in the CMIP5 models. *Journal of Climate 26, 1473–1484.* https://doi: 10.1175/JCLI-D-12-00068.1

Wilkinson, M., Dumontier, M., Aalbersberg, I. & et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data 3, 160018.* https://doi.org/10.1038/sdata.2016.18

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

The Munin Conference on Scholarly Publishing

The short history

Leif Longva

Abstract

The series of annual Munin Conferences was born in 2006. This is a short history of how it happened, and how the conference has developed over the years, step by step, into one of the most important conferences in Europe within its scope. Scholarly publishing is an important part of, and a prerequisite for progress in research. The advent of the Internet has given options for dramatic changes in the process of publishing and the dissemination of research. This has given the Munin Conferences ample issues to debate over the years, and, most likely, the years to come.

In the beginning was Munin

On September 21st 2006 Munin, the open institutional repository of The University of Tromsø was launched with a seminar. At that time, having realized what a powerful tool the Internet was, Higher Education institutions all over the world were launching their open repositories, to disseminate research and student theses.

The development of the Munin repository was governed by a steering committee, led by Library Director Helge Salvesen. The steering committee discussed and decided what should be eligible to include in Munin. Like other institutional repositories, Munin was not meant to be a (formal) publishing channel. It was meant for dissemination of peer reviewed manuscripts, plus student theses. Nevertheless, the steering committee also discussed issues on scholarly publishing in their meetings. And there the idea was born to follow up the launching seminar with another seminar more focusing on publishing issues. The prime issue they wanted to raise was the newly launched system for counting publishing points at the Higher Education institutions, and letting these points to some degree enter the governmental financing model for these institutions.

Thus, the first 'Munin Seminar' was born: held November 23rd 2006 at the UiT campus, and running as an annual seminar/conference every fall since.

'Muninseminaret' was the name the first years. It was held strictly in the Norwegian language as a one day seminar in 2006 and 2007. In

https://doi.org/10.7557/15.5508

2008 and 2009 the seminar was still mainly in Norwegian, but with one or two invited keynotes given in English.

Library Director Helge Salvesen was the Munin Conference initiator, and chairing the Munin Conferences during the first years. © UiT, CC0 1.0 Universal. Personality rights are not licensed.

The Houghton reports and savings from open access

In 2010 the seminar was still a one-day event, but now strictly in English. And the name was changed into The Munin Conference on Scientific Publishing. One of the keynote speakers that year was professor John Houghton from Victoria University Centre for Strategic Economic Studies, Melbourne, Australia. He had done investigations and monetary calculations into the economic impact of open access to scholarly literature in several countries. One example was a calculated 160 million GBP savings and increased research efficiency in the UK from a switch to open access (http://repository.jisc.ac.uk/6476/).

With savings of this magnitude documented, an accelerated switch to open access could be anticipated. However, the traditional model of scholarly publishing has proven to be very resistant to change. The forces involved, and the motivations of the various stakeholders in scholarly publishing have over the years been debated and scrutinized in the Munin Conferences. And after 14 years, the debates have not reached a conclusion yet. There is still a need for the meeting and debating arena of the Munin Conference.

The Munin Conference was never a place purely for open access promotion. Several themes that do not primarily deal with the issue of open or toll access publishing have been raised over the years. Examples are: the peer review process, the publish or perish-pressure (quantity vs. quality in publishing), research data management, the impact factor, the DORA declaration, and pricing models

of licensed access. Nevertheless, the benefits of open and rapid access to research have been demonstrated repeatedly and from various angles at the Munin Conferences.

For those who have followed and attended the Munin Conferences over the years, it is clear that there has been a motion in the focus and in the topics. In the first years much focus was on demonstrating the benefits of open access, and the benefits and strengths of the two roads to open access: green and gold. Over the years, the number of voices within academia questioning the benefits of open access has decreased. So the conference has increasingly focused on the obstacles to open access in terms of stakeholders' agendas and researchers' incentives, financial and budgetary obstacles (including also predatory publishing) as well as various services supporting research and its dissemination. To a large degree open access is a prerequisite for the benefits of such services.

Developing the conference

The model of the Munin Conference was developed further during the planning of the 2011 edition. Until then, the conference was organized with invited speakers only. But now the committee decided to announce a Call for presentations and posters, leaving the final program much depending on the response to this call. Two keynote speakers were invited to set the tone of the conference. And the conference was expanded into a one-and-a-half day event to allow time for a good number of presentations from the open call. This was an exciting and also a daring move – how would the scholarly community respond? Would there be many abstract submissions to the call, or just a barren few? Would the conference and the call be interesting for the international community of scholars with an interest in publishing issues? Tromsø is after all, located far north, and all presenters – except for the invited keynotes – would have to cover their travel costs themselves.

By the deadline of the submission period, the organizing committee could conclude that the move had been successful: the entire one-and-a-half day conference program was filled with exciting presentations. The conference has since moved on to fill two entire days with high quality presentations. From 2011 on, the Munin Conference has also included a session of poster presentations.

Pay-to-publish

As mentioned, the benefits of open access to research have been clearly demonstrated. However, publishing costs do not disappear with open access. The issue of how to finance open access publishing is still being debated heavily – also at the Munin Conferences over the years.

Pay to publish is a common model, at least among the commercial publishers. The shift from paying for access to paying for publishing does imply a new set of concerns. Funding for such a financing model

is not easily available for all; it is a special concern within the humanities, where researchers commonly have limited funds available for such payments. This was a key issue in John Willinsky's keynote talk at the Munin Conference in 2012. John Willinsky is professor at Stanford University School of Education, and is founder of the Public Knowledge Project, which has developed new tools to simplify publishing. Article publishing charge (APC) or book publishing charge (BPC) will not work in the humanities, Willinsky argued.

The funding issue of open access publishing is also very much a concern for research communities in low-income countries. This was a main issue in the keynote talk of Arianna Becerril-García (from Universidad Autónoma del Estado de México) at the Munin Conference in 2019, focusing on scholarly-led, non-profit cooperative publishing initiatives in the Latin America.

It is safe to say that debates about funding models for open access publishing and various associated useful services have not reached a conclusion yet. Here is plenty of questions to be asked and scrutinized at Munin Conferences to come.

An important conference

From 2013 onwards the conference has followed this model of a two-day event with a combination of invited keynote speakers and presentations selected from submitted abstracts from the calls. However, 'scientific publishing' in the name of the annual event has been replaced with 'scholarly publishing'. This change was made following comments that 'scientific' may be understood as covering the STM subject areas only. This was never the intention of the Munin Conference Committee, but rather a nuance of the English language that had slipped the awareness of the committee.

The Munin Conference has truly grown into an important conference on the topic of scholarly publishing. The development in the number of attendants shows this clearly. The first couple of years the seminar attracted approximately 40–50 persons, mainly from the library sector plus Norwegian academic governing bodies and the Research Council of Norway. But already by 2008, the conference had more than 100 attendants, and the number kept rising in the following years – perhaps, not purely because of the northern lights and the wales in the nearby fjords, but also because of the quality of the conference and the timeliness of the topics raised. By 2018 the conference had grown out of the auditorium where it had been located for the preceding ten years, and had to move into the larger auditorium normally used for the big events.

Opening speech for the 13th Munin Conference on Scholarly Publishing 2018. © UiT, CC0 1.0 Universal.

The way the Munin Conference has developed over the years is rather impressive. It is now safe to say that it is one of the most important conferences on scholarly publishing and related topics in Europe. This claim is backed not only by the growth of attendees, but also by the list of keynote speakers the conference has been able to get hold of over the years. (See list of keynote speakers at the end.)

The Munin Conference has always had a goal to attract researchers and graduate students to attend the conference (together with librarians, research administrators, funders as well as publishers) – the researchers are, after all, key stakeholders in the issue of publishing. The number of researchers attending the conference, as well as the proportion of attendants being researchers, has indeed increased over the years, but still less than 10% of attendants are researchers.

The conference has even been free to attend for all UiT-affiliated people, to lower the threshold for UiT's own scholars to attend. Every researcher is, of course, primarily occupied by doing their research, but the questions surrounding publishing and dissemination of their research – as core issues for how research most efficiently may build on previous research – should be of interest to any researcher.

The keynote speakers are the keys

Over the years, the Munin Conference has hosted an impressive list of highly interesting keynote speakers. A few of them are mentioned above. There are lots of others who deserve to be mentioned and praised here, but it would take too long to go through them. Please

see the list of keynote speakers at the end of this paper. The members of the organizing committee have been amazed how positively they all have responded to an invitation to speak, and also how well they have spoken of the conference on their return home. These exciting keynote speakers, and theirs and others' positive feedback to the conference, are always a major motivation factor for taking on the task of preparing the next year's conference.

The list of keynotes is definitely one of the most important success factors of The Munin Conference. The job to shortlist candidates in the preparation of each conference has been governed by a focus on what issue is hot at the time of the actual conference, avoiding to dwell on issues that are more or less out-debated. This has resulted in conferences with timely and exciting content.

The future of the Munin Conference

Will debating the publishing and dissemination processes, and all adjacent issues, ever be exhausted? And will the format of the Munin Conference still develop further?

The answer to the first question is negative, as far as I can see. The question of how to publish research and make it available to peers as well as the interested public is not likely to reach some final conclusion. There are just too many interested parties involved, each with their separate agenda. And there will be external events occurring, influencing the debate on what is the better way to publish and disseminate research. One such external event is the appearance of the corona virus, causing an epidemic that we are in the midst of at the time of writing. The pandemic has made health workers and medics urge for rapid access to the latest research findings on the virus, pinpointing how slow the process of publishing normally is.

Speaking of the corona virus: The pandemic has led to cancellations of numerous events worldwide as a result of infection control regulations, while some events have transformed into online events. And so also for the 15th Munin Conference, scheduled for November 2020. Details are not set at the time of writing, but it is decided that this year's conference primarily will be an online conference.

Whether this shift to an online event will become the normal conference mode for future Munin Conferences remains to be seen. Whichever way it will be, it is likely that UiT's current as well as future library directors may welcome exciting keynote speakers, presenters and attendants for many Munin Conferences to come.

Appendix: A complete list of keynote speakers in The Munin Conference on Scholarly Publishing 2006–2019

Keynote	Institution	Presenting	Year
Fyfe, Aileen	University of St Andrews	Mission or money? Trends in scholarly publishing since c.1850	2019
Jensenius, Alexander Refsum	University of Oslo	Experimenting with Open Research Experiments	2019
Becerril-García, Arianna	University of the State of Mexico	The value of the scholarly-led, non-profit business model to achieve Open Access and scholarly publishing beyond APC: the AmeliCA's cooperative approach	2019
Terry, Robert	World Health Organization's Special Programme for Research and Training in Tropical Diseases (TDR)	Research is born free and everywhere is in chains	2019
Logan, Corina	Max Planck Institute for Evolutionary Anthropology	I won't be #BulliedIntoBadScience	2018
Tennant, Jon	Open Science MOOC	Who should own public science? Preprints, power, and publishers	2018
Peters, Paul	Open Access Scholarly Publishers Association	Ensuring a Healthy and Competitive Market for Open Access Publishing	2018
Tsoukala, Victoria	European Commission	Open Science, Excellent Science! EC policies and initiatives in support of open science	2018
De Rijcke, Sarah	Leiden University	Towards best practices for authorship and research evaluation: Effects of performance metrics and the Leiden Manifesto	2017
Eve, Martin Paul	Birkbeck, University of London	A Matter of Distribution: APC Logic against Consortial Funding Mechanisms	2017
Gowers, Timothy	University of Cambridge	Perverse incentives: how the reward structures of academia	2017

Keynote	Institution	Presenting	Year
		impede scholarly communication and good science	
Curry, Stephen	Imperial College London	Zen and the Art of Academic Maintenance	2016
Kingsley, Danny	University of Cambridge	Reward, reproducibility and recognition in research – the case for going Open	2016
McCabe, Mark J.	Boston University's Questrom School of Management, SKEMA Business School in Sophia Antipolis	How Open Access Affects Competition in Scholarly Publishing Markets: A Tale of Good Intentions, Big Deals and Uncertain Outcomes	2016
Hossenfelder, Sabine	Nordic Institute for Theoretical Physics	Peer review and its discontents	2015
Schekman, Randy	University of California, Berkeley, Howard Hughes Medical Institute	Publishing the most important work in the life sciences	2015
Shieber, Stuart M.	School of Engineering and Applied Sciences, Harvard University	The role of the HE institutions in scholarly publishing and communication	2015
Suber, Peter	Harvard University	What is OA and where is it going – in a conversation on video link with Caroline Sutton	2015
Boulton, Geoffrey	University of Edinburgh/Royal Society UK	Open Data and the Future of Science	2014
Brembs, Björn	Universität Regensburg	When decade-old functionality would be progress – the desolate state of our scholarly infrastructure	2014
Neylon, Cameron	PLOS – Public Library of Science	Managing the Transition to an Open Scholarly Literature	2014
Dechamp, Jean-François	European Commision	The European Commission and Open Access	2013

Keynote	Institution	Presenting	Year
Ferwerda, Eelco	OAPEN Foundation	Open Access in Humanities and Social Sciences	2013
Prosser, David C.	Research Libraries UK	What's 'Open' about Open Access? The Vital Role of Copyright and Licenses	2013
Pattinson, Damian	Public Library of Science	Megajournals and what they mean for the future of scientific publishing	2012
Willinsky, John	Stanford University	The Future of Scholarly Publishing is the Future of Scholarship	2012
Neylon, Cameron	Science and Technology Facilities Council UK	I need to publish more and read less! How new platforms will enable you to publish more effectively while reducing information overload	2011
Rasmussen, Martin	Copernicus Publications	Ensuring Availability and Quality of Research Data through Open Access and Public Peer-Review	2011
Houghton, John	Victoria University's Centre for Strategic Economic Studies (CSES)	Economic impacts of open access in Europe and the United States	2010
Mele, Salvatore	CERN	Open Access Publishing in 2010: what publishers offer, what researchers want – First results of the SOAP project	2010
Heber, Joerg	Nature Publishing Group	Peer review at Nature	2009
Hurum, Jørn	Natural history museum, University of Oslo	Why did we select open access for the publication of IDA? – and did it matter?	2009
Gasser, Urs	University of St Gallen / Harvard Law School	Comparing Open Access Policies of Selected Universities: Towards a Best Practice Model?	2008
Grimson, Jane	Trinity College Dublin	Scientific publication: following up on the recommendations of the European Research Advisory Board	2008

Keynote	Institution	Presenting	Year
Hallén, Arvid	Research Council of Norway	Tilgjengeliggjøring av offentlig finansiert forskning	2007
Vigen, Jens	CERN	Ja takk, begge deler. Om å publisere som før, men nå i fritt leide.	2007
Boye, Erik	Rikshospitalet-Radiumhospitalet, Universitetet i Oslo	Tellekanter, publiseringsmarkedet og konsekvenser for forskerne	2006
Lundberg, Jonas	Medical Management Centre, Karolinska Institutet	Bibliometrics as a research assessment tool – impact beyond the impact factor	2006

RAVNETRYKK

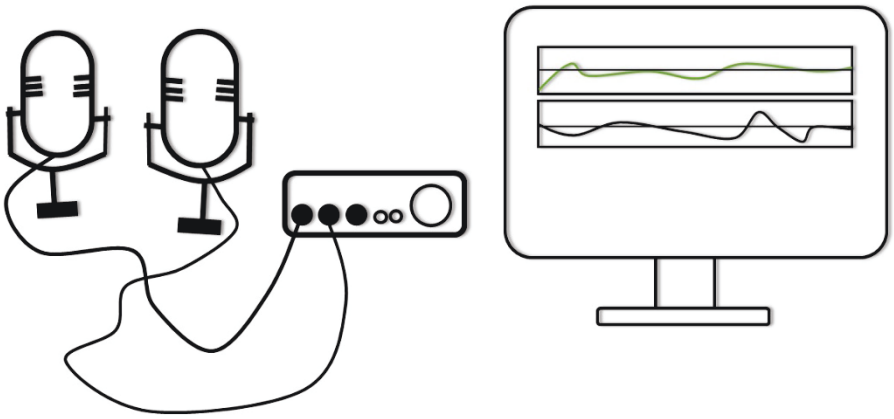
UiT goes open: Et festlig skrift til Stein Høydalsvik

Åpen vitenskap og digital formidling

Erik Lieungh

Sammendrag

I denne artikkelen oppsummerer vi arbeidet med podkasten Open Science Talk, som ble skapt av Universitetsbiblioteket ved UiT Norges arktiske universitet høsten 2018. Dette er en engelskspråklig podkast om åpen vitenskap. Vi forteller litt om motivasjonen bak, hvilke tema vi dekket, hva vi oppnådde, og en kort refleksjon om dette var en god investering for å bedrive digital formidling.



Format og lansering

Tidlig i høstsemesteret 2018 deltok undertegnede på et møte i faggruppe for forsknings- og publiseringsstøtte ved Universitetsbiblioteket i Tromsø. Denne faggruppen, som består av om lag 12 ansatte, har et særlig ansvar for å fremme «åpen vitenskap» (Open Science) i alle sine forgreininger, herunder åpen tilgang til vitenskapelige publikasjoner og forskningsdata. Ideen om en podkast om temaet falt i god jord, og veien fra idé til lansering skulle bli kort.

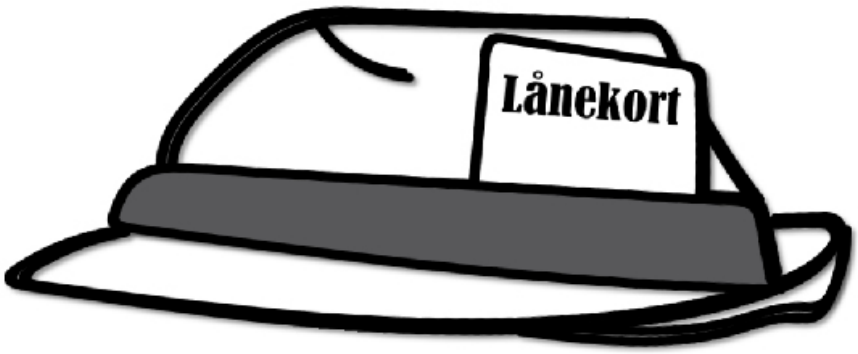
Det var et klart dogme i hvordan podkasten ble laget. Hver episode skulle være cirka 15 minutter lang. Den skulle ikke være uttømmende i natur, men presis og fokusert for det temaet for episoden handlet om. Det var en programleder i hver episode samt en gjest. Podkasten var på engelsk, noe vi i forkant mente det kunne komme kritikk av, men som det aldri gjorde. Årsaken var rett og slett

https://doi.org/10.7557/15.5495

den store andelen forskere og ph.d.-studenter ved UiT som er utenlandske. Dessuten innebar det at vi kunne sende podkasten ut i verden, hvor det finnes mange i samme situasjon som oss selv, som også hadde et behov for å informere om åpen vitenskap.

De fire første episodene kom på lufta i rask rekkefølge i månedsskiftet september/oktober 2018. Førstemann ut var Stein Høydalsvik, som laget introduksjonsepisoden om selve begrepet «åpen vitenskap»[1]. Så fulgte Jan Erik Frantsvåg med sin episode om den nylig lanserte Plan S[2], deretter kom Aysa Ekanger og fortalte om hvilke innvendinger som var mot «åpen tilgang», og hvilke som var mer rasjonelle enn andre[3]. Deretter var det Lars Figenschou med sin episode om hvorfor det var viktig å sikre forskningsdata fra eldre forskere[4].

Dette var en sterk start for en podkast, og satte standarden for det som skulle komme seinere.



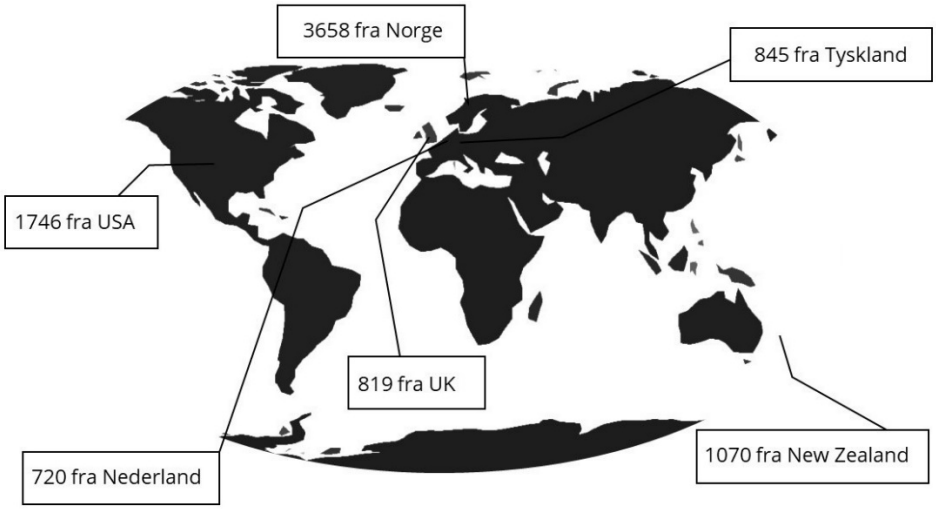
Stor bredde

I skrivende stund er det laget 33 episoder. Mange kollegaer ved Universitetsbiblioteket, fra andre enheter ved UiT Norges arktiske universitet, og fra utgivere, andre bibliotek og universitet, har deltatt. Alle delte mer enn gjerne av sin kunnskap, ut fra sine fagområder og erfaringer.

Vi har også kunnet bringe til torgs helt nye perspektiv, godt hjulpet av fri tilgang til gjestene under de årlige Munin-konferansene. Rett etter nå avdøde Jon Tennants flammende innlegg på talerstolen i

[1] Høydalsvik, S., & Lieungh, E. (2020). What is Open Science? Open Science Talk, (1). https://doi.org/10.7557/19.5266
[2] Frantsvåg, J. E., & Lieungh, E. (2020). What is Plan S? Open Science Talk, (2). https://doi.org/10.7557/19.5267
[3] Ekanger, A., & Lieungh, E. (2020). To OA or Not to OA? Open Science Talk, (4). https://doi.org/10.7557/19.5287
[4] Figenschou, L., & Lieungh, E. (2020). Senior Scientists & Valuable Data. Open Science Talk, (3). https://doi.org/10.7557/19.5286

Aud 1, gikk han over i studio og forklarte nærmere om sin noe aktivistiske tilnærming til åpen vitenskap[5]. Munin-konferansene har også gitt oss anledning til å hanke inn spennende perspektiv i form av intervjuer om åpen vitenskap på den sørlige halvkule[6], i Midtøsten, i Latin-Amerika[7]. Vi har fått høre om mange ulike initiativer fra England, Nederland og Tyskland. Intervjuobjekter har fortalt om ulik bruk av åpen vitenskap i mange disipliner: kjemi, musikk, historie, psykologi, språkvitenskap, for å nevne noen. Vi vet også av de cirka 11.000 avspillingene, at vi har lyttere fra USA, New Zealand, Nederland, England, Tyskland og Sverige, for å nevne de største.



Statistikk over avspillinger av Open Science Talk fordelt på land, innhentet 27. april 2020. Kun de seks landene med flest avspillinger vises her.

Digital formidling

Men var det hele en suksess? Det er vanskelig å si. Selv har jeg for et par måneder siden sluttet ved Universitetsbiblioteket for å begynne i en annen jobb. Selv om jeg personlig ikke lenger er en del av kollegiet, våger jeg å påstå at for Universitetsbiblioteket har det vært en suksess. Vi har skapt et produkt sammen for en liten pengesum, vi har profilert UiT som et «åpen vitenskap-universitet», og det er mange der ute som kjenner Open Science Talk.

Vi vet også at da vi trengte den som mest, så var podkasten meget aktuell. Vi kunne lage en episode om Plan S, kort tid etter at planen var lansert. Vi kunne fortelle om Norges brudd med forlaget Elsevier

[5] Tennant, J., & Lieungh, E. (2020). The Future of Open Science. Open Science Talk, (11). https://doi.org/10.7557/19.5294

[6] Hachani, S., & Lieungh, E. (2020). Publishing in the Global South. Open Science Talk, (22). https://doi.org/10.7557/19.5305

[7] Becerril-García, A., & Lieungh, E. (2020). Open Access in Latin America. Open Science Talk, (24). https://doi.org/10.7557/19.5307

to dager etter avtalen var brutt[8], og følge opp umiddelbart når det kom på plass en ny avtale[9]. Vi hadde en egen plattform til å fortelle om svenskenes nye avtale med samme forlag, kun en uke etter at de hadde inngått den. Gjesten var direktør for Stockholms universitets-bibliotek, Wilhelm Widmark[10]. På dette tidspunktet hadde få andre medier omtalt avtalen.

Men det finnes en stor fallgruve når man lager digital formidling, spesielt når man bruker sosiale medier som kanal. Det er at man hele tiden står i fare for å kun formidle til sin egen menighet. Algoritmene er sterke, og ekkokamrene mange. Det gjør at de som dultet borti en åpen vitenskaps-podkast fra Tromsø, nok oftest selv allerede var frelst. Vi får bare håpe at de mange digitale hjertene og tomlene fra entusiastene lot seg spre til også andre kamre av den digitale verdenen, men vi kan nok ikke si at selve podkasten i seg selv stimulerte til særlig debatt. Det er verdt å legge merke til, ettersom temaet på ingen måte er ukontroversielt.

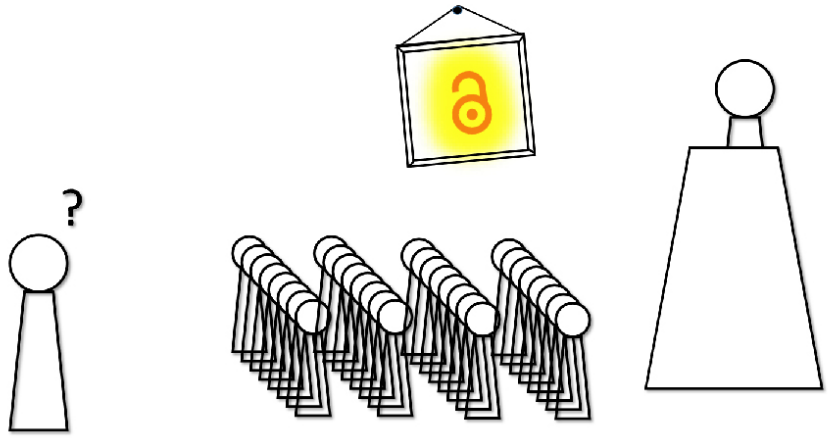
Foruten tallene, kan vi også finne igjen spor etter podkasten rundt omkring i verden. Fra universitet og bibliotek som lenker den opp på sine egne sider, til internsidene fra en av de største utgiverne i verden, Wiley, til flere forskeres digitale publiserings-CVer.

Episodene er nå også blitt arkivert på en fornuftig måte gjennom Universitetsbibliotekets egen publiseringstjeneste, Septentrio Academic Publishing. Det innebærer at for framtiden vil episodene alltid ligge der enten som en ressurs, eller som et historisk arkiv for hva som rørte seg i åpen vitenskaps-miljøet disse årene.

[8] Magnussen, M., & Lieungh, E. (2020). No Deal with Norway. Open Science Talk, (15). https://doi.org/10.7557/19.5298

[9] Magnussen, M., & Lieungh, E. (2020). Norway made a new deal with Elsevier. Open Science Talk, (17). https://doi.org/10.7557/19.5300

[10] Widmark, W., & Lieungh, E. (2020). Sweden made a new deal. Open Science Talk, (23). https://doi.org/10.7557/19.5306



Avslutning

Om podkastserien i seg selv har skapt noe større forståelse for hva åpen vitenskap er, er usikkert. Det som er sikkert, er at det i hvert fall er informasjon tilgjengelig for alle der ute. Åpen vitenskap er ikke i synsfeltet til vår egen presse. Det er ikke noe kritikkverdig i det, men det er rett og slett ikke noe som opptar innbyggerne i landet utenfor UH-sektoren. I ny og ne drypper det litt åpen vitenskap ned mellom sidene i avishusenes trykkerier, men det er gjerne da diskusjonen i fagbladene har vært på sitt kvasseste. For eksempel når godt etablerte professorer skyter fra hoftene fordi bibliotekariatet, for en gangs skyld, hever sin egen stemme. Men der det skrives om åpen vitenskap er det som regel forbundet med konflikt, og den evige kampen mellom det som alltid har vært og det som kommer nå.

Poenget er uansett: Ingen kan si at de ikke har tilgang til kunnskapen om åpen vitenskap. For der har universitetsbiblioteket i Tromsø bidratt, med sterke fagfolk som byr på seg selv, uvante og nye perspektiv, og med et oppriktig ønske om å dele kunnskap. Det er disse ingrediensene som skal til i kampen om oppmerksomheten, som må ligge til grunn for at formidlingen skal bryte gjennom muren. Open Science Talk er i hvert fall et ærlig forsøk på det siste, et forsøk det er verdt å være stolt over.

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

Hva med åpen læring?

Mariann Løkse

Sammendrag

Nye erfaringer kan tvinge fram nye tanker og i noen tilfeller ny praksis. Våren 2020 ble et unikt semester på mange måter. Fra en dag til den neste ble digital undervisning normen og ordinære forelesninger og seminarer ble satt på pause. Kan vi bruke noen av disse erfaringene til å tenke nytt om deling av læringsressurser? Universitetsbiblioteket ønsker å knytte erfaringer fra det mangeårige arbeidet med åpen tilgang til forskning til de erfaringene som nå gjøres med digitalisering av undervisning. Er det på tide å tenke åpen undervisning på samme måte som vi tenker åpen forskning? Til tross for vesensforskjeller i både teori, metoder og praksis, ser vi også likheter som kan være med på å bane vei for en mer helhetlig tankegang omkring åpenhet i akademia.

Innledning

Vi har kommet langt i arbeidet med å gjøre åpen vitenskap til normalen ved UiT, selv om vi ennå ikke er i mål. Med et semester preget av korona, med stengte biblioteker og stort behov for tilgang til ny forskning, har spørsmålet om å gjøre offentlig finansiert forskning åpent tilgjengelig for alle blitt mer relevant enn noensinne. Samtidig har et annet spørsmål relatert til åpenhet blitt aktualisert, nå som svært mye av undervisningen gjøres digital. Er det på tide å tenke at også undervisningsressurser skal være åpne? Er det rett at videoforelesninger og nettkurs låses inn på læringsplattformer som kun noen få har tilgang til, eller skal også disse ressursene kunne komme alle til gode?

Åpne læringsressurser

Etter mange år med fokus på åpen tilgang og åpen vitenskap, der målet er at mest mulig av publisert forskning, og tilhørende forskningsdata, gjøres åpent tilgjengelig for alle, rettes nå blikket mot undervisningsressurser også. Skal det være like selvsagt å dele digitale læremidler som forskningsresultater?

Åpne digitale læringsressurser, bedre kjent som OER (*open educational resources*), er digitale læremidler som er åpent lisensiert. Det betyr at de kan tas fritt i bruk, deles og endres av andre.

https://doi.org/10.7557/15.5501

Argumentasjonen her er i stor grad den samme som for forskningsresultater. Offentlig finansiert utvikling av læremidler bør komme det offentlige til gode, og ikke legges inn i lukkede læringsplattformer eller institusjonelle nettsteder. UNESCO har brukt termen OER siden 2002 og definerer det som:

> "teaching, learning and research materials in any medium – digital or otherwise – that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions." (UNESCO, 2019)

Hovedmålet er økt tilgang til kunnskap og læringsmateriale for alle, og dermed en tilrettelegging for kompetanseheving i samfunnet, uavhengig av økonomi og hvilken del av verden man tilhører, eller om man har mulighet til å studere ved en tradisjonell utdanningsinstitusjon eller ikke. For de som ønsker å lære (elever og studenter) betyr OER gratis læremidler, tilgjengelig digitalt når man trenger dem. For de som underviser betyr det at man kan spare tid og penger ved å bruke materiale som allerede finnes, man kan tilpasse det til sine egne studenter og egne læringsmål, og eventuelt videreutvikle læremidlene ytterligere.

Livslang læring

I rapporten *Lærekraftig utvikling: Livslang læring for omstilling og konkurranseevne* vektlegges behovet for å tilrettelegge for både etter- og videreutdanning og generell kompetanseheving. Dette behovet gjelder ikke noen få næringer som må omstilles for å møte framtida, men de aller fleste yrker. For å sitere fra rapporten: det er ikke lenger «tilstrekkelig å utdanne seg 'en gang for alle'» (NOU 2019:12, 2019, s. 13).

For de aller fleste yrkesaktive vil altså behovet for å lære mer eller lære noe nytt oppstå i løpet av karrieren. Mange vil også skifte karriere, enten av nødvendighet, eller fordi man oppdager nye og interessante utfordringer innen andre bransjer. Felles for alle er at behovene og tilbudene er forskjellige og avhenger blant annet av økonomi, geografi og bakgrunn.

Moderne teknologi gir oss muligheten til å være studenter hjemme i vår egen stue. Vi kan kommunisere med andre studenter og lærere, vi har tilgang til digital forskningslitteratur, og vi kan dele erfaringer, tanker og ideer online. Det er selvsagt ikke gitt at nettbasert læring bare er positivt. Ofte går man glipp av lærerike samtaler og sosiale fordeler ved å være en del av et akademisk fellesskap, men også her finnes det nå gode verktøy for å samhandle bedre digitalt.

Nettbasert læring og utdanning betyr at arbeidstakere som ønsker å lære litt mer, eller fordype seg i et emne, kan gjøre dette helt på egen hånd. I mange bygder og tettsteder er man avhengig av at folketallet holder seg stabilt og at bedriftene på stedet er i stand til å levere tjenester det er behov for. I bransjer med stor konkurranse vil det gi

store fordeler med kompetente og oppdaterte ansatte. Både innen helse og omsorg, skolevesen og industri skjer det en konstant utvikling som gjør det nødvendig med lett tilgjengelig kunnskap.

Åpne læringsressurser skaper en type fleksibilitet som gjør at kompetansebygging kan skje overalt i landet, ikke bare der det finnes store universitets- eller høgskolecampus, en faktor som kanskje er spesielt viktig i et grisgrendt land som vårt. Det gjør også at arbeidstakere kan oppdatere seg på faget sitt når det passer dem, og dermed slippe å måtte ta permisjon fra eller slutte i jobben for å flytte til nærmeste studiested. For mange er dette en motivasjonsfaktor som gjør det langt enklere å ta steget inn i etter- og videreutdanning.

Læring i et kunnskapssamfunn – eksemplet iKomp

En viktig faktor ved kunnskapssamfunnet er at kunnskapen er åpent tilgjengelig, gratis for alle. Men en nesten like viktig side ved dette er å ha gode strategier for å velge riktige informasjonskilder. Dersom både læringsressurser og forskningen serveres som et kunnskapens overflødighetshorn, hvordan skal man da vite hvor man skal starte, hva man skal stole på og hva man kan forkaste?

Gjennom mange år har samtaler og diskusjoner om åpen tilgang preget hverdagen på Universitetsbiblioteket. Vi som hadde undervisning som spesialområde hadde riktignok oppmerksomheten rettet mot andre oppgaver enn å bygge infrastruktur og holdningsarbeid for åpen tilgang, men da vi satte oss ned for å lage et nettkurs om kildebruk og læringsstrategier føltes det helt naturlig å velge å publisere kurset på en åpen plattform. At valget falt på plattformen OpenEdX var helt bevisst fra vår side, og i ettertid ser vi at våre argumenter i stor grad var preget av de samme tankene som ligger bak open access-bevegelsen.

Informasjonskompetanse handler i stor grad om å gjøre studenter og befolkningen for øvrig i stand til å fatte beslutninger og foreta valg basert på kunnskap. Det handler om å øke forståelsen av den informasjonen vi omgir oss med og å skape ny kunnskap basert på forskning og vitenskapelige metoder. Dette anser vi som en grunnleggende og selvsagt kompetanse for alle studenter og forskere innen høyere utdanning.

Et universitet eksisterer imidlertid ikke separat fra det øvrige samfunnet, og fra vårt ståsted ved Universitetsbiblioteket ser vi det som en viktig oppgave å bidra til økt kompetanse også blant ungdom og voksne for øvrig, både regionalt og nasjonalt. Dette er i tråd med UiTs strategi, *Drivkraft i nord*, som slår fast at «UiT skal bidra til at forskningsbasert kunnskap ligger til grunn for samfunnsdebatten» (UiT, 2019).

I tillegg er det et viktig prinsipp for bibliotekbransjen å ha et tilbud som fremmer åpenhet, deling og gratis tilgang til kunnskap. Prinsippet om åpenhet er forankret i både UiTs strategi og i Universitets- og høyskoleloven. Sistnevnte fremholder at

> «Universiteter og høyskoler skal sørge for åpenhet om resultater fra forskning eller faglig eller kunstnerisk utviklingsarbeid». (Universitets- og høyskoleloven, 2005, §1-5)

Vi ønsket også å unngå problematikken at de samme produktene må utvikles på nytt ved hver institusjon fordi så mye ligger innelukket i en læringsplattform eller krever annen form for institusjonell tilhørighet. Vi ønsket å lage en læringsressurs som alle kan dra nytte av, uavhengig av hvem de er eller hvor de bor. Vi valgte derfor å legge nettkurset – som etter hvert fikk navnet *iKomp* – på en åpen plattform, slik at hvem som helst kan ta kurset, eller bruke kurset som utgangspunkt og inspirasjon, og kanskje lage noe enda bedre selv. Det gir oss anledning til å utveksle råd og tips og dele erfaringer og tilbakemeldinger med andre som jobber på samme måte som oss. Da vi gjorde grunnarbeidet til kurset benyttet vi oss selv for eksempel av andre åpne kurs, både om informasjonskompetanse og beslektede emner innen læring og kildebruk.

Som tema er informasjonskompetanse viktigere enn noensinne. Det er et viktig verktøy i bekjempelsen av pseudovitenskap og falske nyheter. I kampen mot det (bevisst eller ubevisst) feilinformerte samfunn har alle kunnskapsinstitusjoner et stort ansvar for å tilby fakta, forskningsbasert kunnskap og kompetanse til alle som deltar eller følger med på samfunnsdebatten. Dette er viktig for at vi som samfunn skal utvikle oss videre, og det er viktig for enkeltindividers valg i forhold til blant annet helse, utdanning, yrkesliv og politisk deltakelse.

I vår definisjon av informasjonskompetanse finner vi også kompetanse om læring i seg selv, altså strategier for å lære bedre, for dermed å kunne dra bedre nytte av informasjonen og kunnskapen som finnes. Dette fant vi nødvendig å inkludere etter å ha lyttet til de mange studentene vi har snakket med opp gjennom årene. I overgangen fra videregående til universitet skjer det mange forandringer i både undervisningsmetoder og lærestoff. Studenter forventes i større grad enn elever i videregående skole å selv sette seg inn i faglitteraturen, følge undervisning og avgjøre hva som er viktig og hva som er mindre viktig. For noen krever dette nye læringsstrategier. For andre krever det en større dose selvdisiplin enn de er vant med fra tidligere. Da er det nyttig med enkle tips om hvordan studiehverdagen kan gjøres mer strukturert og håndterbar.

Det åpne nettkurset iKomp ligger gratis tilgjengelig på nett for alle som er interessert i å bruke det. Det samme gjelder søsterkurset *iKomp VGS*, utviklet av UiT i samarbeid med Troms fylkeskommune. Ved å opprette et kurs for skoleelever tar vi vårt samfunnsoppdrag som akademisk institusjon på alvor og prøver å bidra til en bedre forståelse og bruk av kunnskap, i tillegg til å gjøre elevene bedre rustet til å møte høyere utdanning og arbeidsliv.

Veien videre

Mange spørsmål knyttet til for eksempel kvalitet og opphavsrett gjør at fokuset på åpne læringsressurser ennå er i en tidlig fase. Vi som har fulgt utviklingen av åpen tilgang til forskning og forskningsdata gjennom flere år, kjenner igjen alle argumentene og all usikkerheten knyttet til denne nye måten å tenke på. Og vi vet at disse spørsmålene ennå ikke er løst. Derfor tror vi at produksjon av åpne læringsressurser er en ide som trenger modning og som trenger tid, slik at vi ivaretar både studenters, underviseres og utdanningsinstitusjoners behov. Men vi tror også at tiden er inne til å sette fokus på dette temaet i mye større grad enn tidligere. La oss starte nå.

Referanser

iKomp.no. https://result.uit.no/ikomp

NOU 2019:12 (2019). Lærekraftig utvikling: Livslang læring for omstilling og konkurranseevne. Oslo: Kunnskapsdepartementet.

UiT Norges arktiske universitet (2019). Drivkraft i nord. Hentet fra https://uit.no/om/art?p_document_id=355830&dim=179033

UNESCO (2019). Open Educational Resources (OER). Hentet fra https://en.unesco.org/themes/building-knowledge-societies/oer

Universitets- og høyskoleloven – uhl. Lov om universiteter og høyskoler av 1. april 2005 nr. 15. Hentet fra https://lovdata.no/dokument/NL/lov/2005-04-01-15?q=uhl

RAVNETRYKK

UiT goes open: Et festlig skrift til Stein Høydalsvik

How we tried to JATS XML

Obiajulu Odu and Aysa Ekanger

Abstract

This is a story about how an Open Journal Systems-based library publishing service tried (and failed) to implement XML in one of its publications. We ran a small project to look at how journals we support could develop a JATS XML-based publishing workflow using existing open software tools.

Introduction

Septentrio Academic Publishing is a library-run service at UiT The Arctic University of Norway that publishes open access journals on the Open Journal Systems (OJS) platform.[1] Septentrio has always strived to optimize its journals for the digital format, by looking at issues such as screen-friendly PDF layout, long-term archiving, and visibility in the open access infrastructure through journal indexing and permanent identifiers.

If you only publish journal articles in human-readable formats, like PDF, you are likely to be missing valuable indexing opportunities. JATS-compliant XML is increasingly the standard indexing format. Publishing articles in XML format has long been a discussion in the Septentrio team because of XML's benefits and advantages. However, limited resources of the service have prevented the team from bringing the plans to fruition.

When the University Library decided to publish an issue of the Septentrio-series *Ravnetrykk* dedicated to Stein Høydalsvik, the Septentrio team thought: how about publishing this issue in the XML format? Stein coordinated the Septentrio service for a number of years and it would be a nice tribute to Stein if Septentrio's first XML experience was an issue dedicated to him.

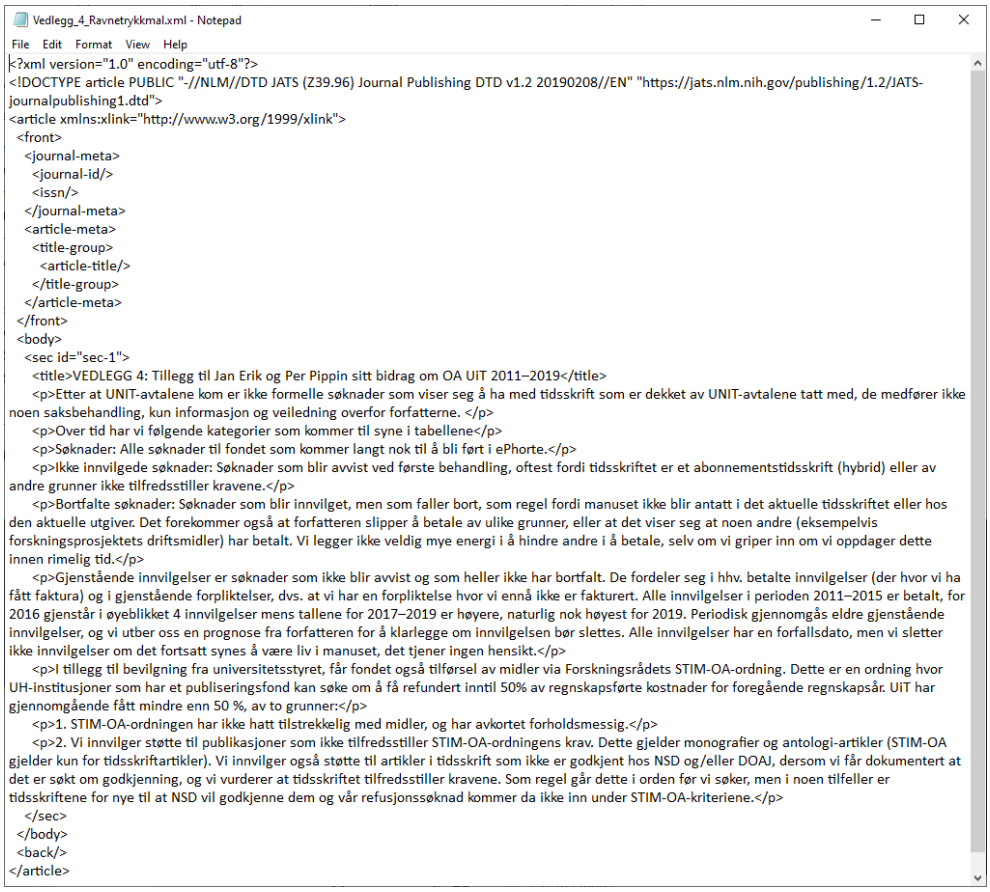
[1] OJS was released by the Public Knowledge Project (PKP) in 2001 and is, according to PKP, "the most widely used open source journal publishing platform in existence" (https://pkp.sfu.ca/ojs/).

https://doi.org/10.7557/15.5517

In this short article we give an overview of the work that was done to publish an issue in *Ravnetrykk* using JATS XML format. We explain the choices made, challenges met and experiences gained.

Reviewing of JATS tools for XML Publishing

Extensible Markup Language (XML) is a markup language that allows to encode content in a machine-readable and layout-independent way, which is more flexible and reusable for a variety of formats (PDF, HTML, etc.). XML makes documents more searchable and accessible, and better-suited for preservation; it allows text mining and opens for content enrichment through multimedia and semantic tagging.[2] The XML format has long been a discussion in the Septentrio team because of its benefits.


Vedlegg_4_Ravnetrykkmal.xml - Notepad

File Edit Format View Help

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE article PUBLIC "-//NLM//DTD JATS (Z39.96) Journal Publishing DTD v1.2 20190208//EN" "https://jats.nlm.nih.gov/publishing/1.2/JATS-
journalpublishing1.dtd">
<article xmlns:xlink="http://www.w3.org/1999/xlink">
 <front>
  <journal-meta>
   <journal-id/>
   <issn/>
  </journal-meta>
  <article-meta>
   <title-group>
    <article-title/>
   </title-group>
  </article-meta>
 </front>
 <body>
  <sec id="sec-1">
   <title>VEDLEGG 4: Tillegg til Jan Erik og Per Pippin sitt bidrag om OA UiT 2011–2019</title>
   <p>Etter at UNIT-avtalene kom er ikke formelle søknader som viser seg å ha med tidsskrift som er dekket av UNIT-avtalene tatt med, de medfører ikke
noen saksbehandling, kun informasjon og veiledning overfor forfatterne. </p>
   <p>Over tid har vi følgende kategorier som kommer til syne i tabellene</p>
   <p>Søknader: Alle søknader til fondet som kommer langt nok til å bli ført i ePhorte.</p>
   <p>Ikke innvilgede søknader: Søknader som blir avvist ved første behandling, oftest fordi tidsskriftet er et abonnementstidsskrift (hybrid) eller av
andre grunner ikke tilfredsstiller kravene.</p>
   <p>Bortfalte søknader: Søknader som blir innvilget, men som faller bort, som regel fordi manuset ikke blir antatt i det aktuelle tidsskriftet eller hos
den aktuelle utgiver. Det forekommer også at forfatteren slipper å betale av ulike grunner, eller at det viser seg at noen andre (eksempelvis
forskningsprosjektets driftsmidler) har betalt. Vi legger ikke veldig mye energi i å hindre andre i å betale, selv om vi griper inn om vi oppdager dette
innen rimelig tid.</p>
   <p>Gjenstående innvilgelser er søknader som ikke blir avvist og som heller ikke har bortfalt. De fordeler seg i hhv. betalte innvilgelser (der hvor vi ha
fått faktura) og i gjenstående forpliktelser, dvs. at vi har en forpliktelse hvor vi ennå ikke er fakturert. Alle innvilgelser i perioden 2011–2015 er betalt, for
2016 gjenstår i øyeblikket 4 innvilgelser mens tallene for 2017–2019 er høyere, naturlig nok høyest for 2019. Periodisk gjennomgås eldre gjenstående
innvilgelser, og vi utber oss en prognose fra forfatteren for å klarlegge om innvilgelsen bør slettes. Alle innvilgelser har en forfallsdato, men vi sletter
ikke innvilgelser om det fortsatt synes å være liv i manuset, det tjener ingen hensikt.</p>
   <p>I tillegg til bevilgning fra universitetsstyret, får fondet også tilførsel av midler via Forskningsrådets STIM-OA-ordning. Dette er en ordning hvor
UH-institusjoner som har et publiseringsfond kan søke om å få refundert inntil 50% av regnskapsførte kostnader for foregående regnskapsår. UiT har
gjennomgående fått mindre enn 50 %, av to grunner:</p>
   <p>1. STIM-OA-ordningen har ikke hatt tilstrekkelig med midler, og har avkortet forholdsmessig.</p>
   <p>2. Vi innvilger støtte til publikasjoner som ikke tilfredsstiller STIM-OA-ordningens krav. Dette gjelder monografier og antologi-artikler (STIM-OA
gjelder kun for tidsskriftartikler). Vi innvilger også støtte til artikler i tidsskrift som ikke er godkjent hos NSD og/eller DOAJ, dersom vi får dokumentert at
det er søkt om godkjenning, og vi vurderer at tidsskriftet tilfredsstiller kravene. Som regel går dette i orden før vi søker, men i noen tilfeller er
tidsskriftene for nye til at NSD vil godkjenne dem og vår refusjonssøknad kommer da ikke inn under STIM-OA-kriteriene.</p>
  </sec>
 </body>
 <back/>
</article>
```

Figure 1. Example of JATS XML

There are a lot of XML-producing tools – we need tools compatible with OJS, and namely those that produce XML in JATS format. The Journal Article Tag Suite (JATS) is an XML tag set format to

[2] See more at https://en.wikipedia.org/wiki/XML. An overview of the benefits of XML can be found in a blog post by Shanu Kumar, *JATS XML: Everything a Publisher Needs to Know*, https://blog.typeset.io/jats-xml-everything-a-publisher-needs-to-know-95862a4184a3.

structure, share and archive academic articles, that has become an international standard.[3] The tools should also be open source: proprietary solutions were intentionally excluded in our project because of Septentrio's overall commitment to openness and freedom of knowledge. It was also preferable to choose a tool that converts DOCX (Word) files to XML, as all authors in the *Ravnetrykk* issue (and almost all authors of the rest of Septentrio journals) submit their manuscripts as DOCX files.

A good starting point for our project were tools presented at the XML Publishing Workflow workshop that took place at the PKP 2019 International Scholarly Publishing Conference.[4]

The picture below depicts the workflow for JATS XML in Septentrio. We divided the JATS workflow in four stages: submission, conversion, editing and presentation/visualization stages. XML can be introduced at different stages of the production process, with XML-first, XML-last and XML-middle workflows.

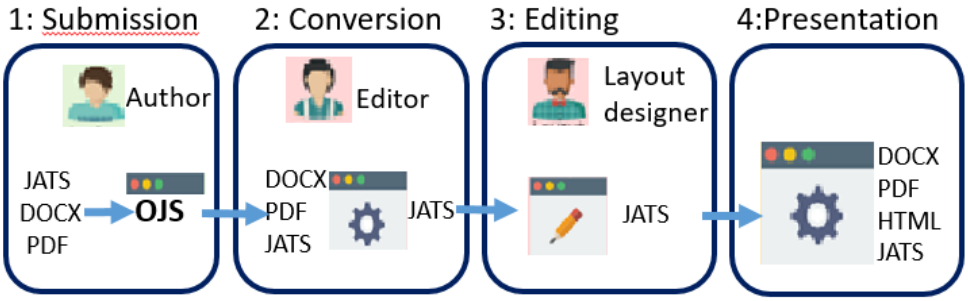


Figure 2. Stages of JATS XML workflow

Making Choices in Different Stages

Our search for a free software tool that is robust, perdurable and standard-compliant, consisted of soliciting help in the PKP Community Forum and testing tools on GitHub compatible with OJS 3.1/3.2 or later. We excluded tools that did not show any sign of development in the last two years or seemed abandoned.

Submission stage

As mentioned above, we decided to focus on XML tools that work with the DOCX format as this format is used by the authors of most of Septentrio's journals and series.

Septentrio also has two journals where the preferred format of submissions is LaTeX; if an XML workflow is to be developed for all of Septentrio's journals in the future, we need to consider tools that can work with LaTeX.

---

[3] https://en.wikipedia.org/wiki/Journal_Article_Tag_Suite

[4] We briefly discussed the workshop in Eikebrokk et al. (2019), https://doi.org/10.7557/11.5204.

Conversion stage

The originals (DOCX-files) from authors are transformed to JATS XML. We did a review of two tools that work with DOCX files: meTypeset and docxToJats.

meTypeset[5]: is a standalone command-line tool written in Python to convert from DOCX format to JATS XML. It requires the text to be formatted in a particular way and gives mixed results. Last changes in the development seemed to have taken place more than a year before we started reviewing the tool.

docxToJats[6]: The tool works as a standalone application written in PHP that allows making batch transformations of articles from DOCX to JATS XML. The tool, which is being actively developed, is still a «work in progress», so a lot of manual cleaning of the output XML file is needed. docxToJats and its OJS plugin, JATS Parser, seemed the most appropriate for our use case as it has the most recent developments.

Editing stage

The docxToJats tool is a «work in progress», not all features usually present in a manuscript (e.g. figures, formulas and footnotes) are supported. If a conversion is not perfect, or some modification must be made in the final galley, a tool is necessary to make changes directly in the produced JATS XML files.

For this stage of the workflow, a relevant tool is Texture[7] – a visual JATS XML editor that allows authors/editors to produce documents in JATS from scratch or edit existing files. It has a simple, user-friendly graphical interface and it is available as a standalone tool and as an OJS plugin. Note, however, that it supports only a strict subset of JATS elements, which some may find too restrictive. A thorough test of the tool for this purpose is pending.

Presentation stage

JATS XML needs to be converted to something more human-readable, like HTML or PDF. We tested two OJS3 plugins: Lens Galley viewer plugin and JATS Parser plugin.

Lens Galley viewer plugin[8] integrates eLife Lens for OJS 3.0. Lens Galley viewer is a well-developed tool that displays the full text in a separate window, with flexible navigation options. However, it is not optimized for display on mobile phones and tablets. It did not work

[5] https://github.com/MartinPaulEve/meTypeset
[6] https://github.com/Vitaliy-1/docxToJats
[7] https://github.com/pkp/texture
[8] https://github.com/asmecher/lensGalley)

in OJS 3.1 and 3.2, and it collides with Texture plugin (as pointed out in the PKP Community Forum).

The JATS Parser plugin[9] parses JATS XML and displays it on the article's abstract page in OJS (the landing page for the article's DOI) as HTML. The plugin also allows to opt for auto-generation of a PDF from the XML: a PDF link is then shown on the article abstract page. The plugin parses JATS content (<body> and <back> sections), whereas metadata, such as title, author and article abstract, are extracted from the submission metadata in OJS. Elements such as figures, formulas and footnotes still need to be implemented in the plugin. JATS Parser does not work in all OJS themes, and works on OJS 3.2.x, but not on OJS 3.1.x (which Septentrio currently is based on). Nevertheless, despite these shortcomings, JATS Parser at the moment is the best solution for the Presentation stage.

Challenges and Experiences

To-do lists at our library department are usually long, and the spring of 2020, with its COVID-19 disruptions, did not contribute to making them shorter or easier to move through – rather the opposite. The editor-moderated final versions of the manuscripts started arriving by the beginning of May 2020 and it was about that time that the exploration of XML tools started. The team have not achieved fully the implementation in the mentioned four stages due to shortage of time and the limited resources of the service. Nevertheless, this was a useful learning experience for us, and potentially useful for others who intend to implement the workflow.

docxToJats did not support all major features of DOCX.[10] The planned 1.0.0 release will likely be the first stable release.

When a DOCX file is processed, the text is converted into an XML file, whereas non-textual elements like images, embedded Excel objects, tables and figures are extracted into attached files. These attached files must be uploaded as dependent files to XML galleys in OJS.

[9] https://github.com/Vitaliy-1/JATSParserPlugin

[10] An overview of the elements that are already supported and are planned to be developed in the near future is available at https://github.com/Vitaliy-1/docxConverter#what-article-elements-are-supported.

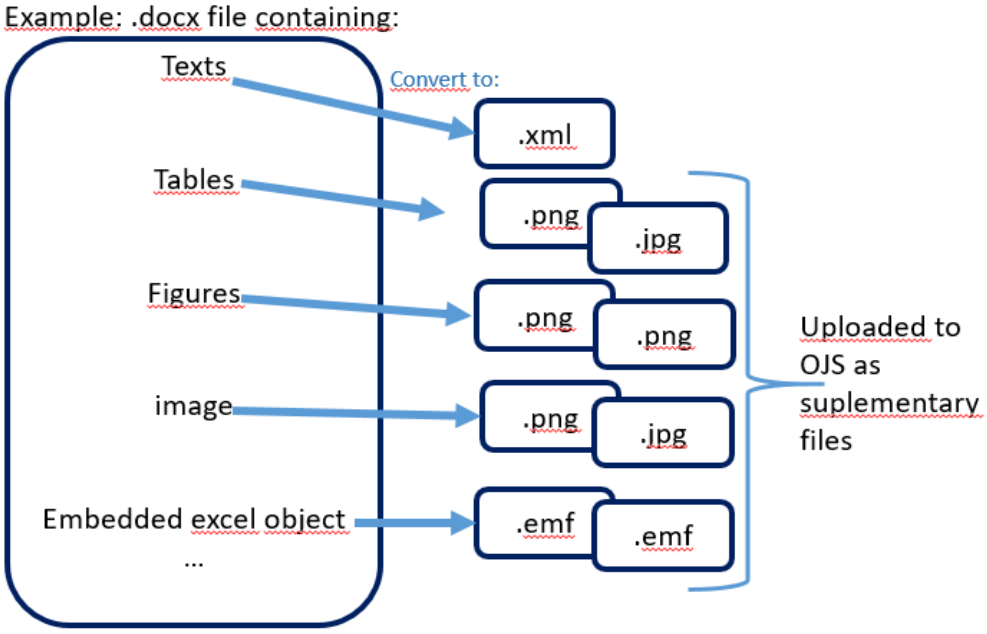


Figure 3. From a DOCX file to an output XML file and attached files

Each dependent file in OJS has associated fields for captions and rights information. Table and figure captions functionality in the docxToJats plugin is planned for the 1.0.0 release; it will be interesting to see whether there will be an option in the JATS Parser plugin to choose where the captions information must be imported from – the XML file or dependent file metadata.

Septentrio is currently running on OJS 3.1.2.0. We started testing XML conversion in a test environment running 3.2.0.2, which happened to be a newer version of OJS. The initial results were encouraging: the articles (consisting of text and images) were converting and displaying nicely. There were a few issues with the non-textual elements not supported by the plugin: but we could find temporary workarounds. We decided to continue in the Septentrio production server and work out the possible remaining issues there. To our disappointment, however, the JATS Parser plugin did not work in OJS 3.1.2.0, and we did not find any evidence that the plugin works or has been tested on OJS 3.1.2.0. It was now a little over one week before the issue was supposed to be published. If we wanted to present Stein with an issue in the XML format, we had to upgrade our entire OJS installation, with all the 19 journals and series. An update at such short notice, however, would probably be stressful for the editors of other Septentrio journals, who would have to put up with at least a day of downtime they had not planned for. Another complicating factor was that we lacked a complete Norwegian translation for the new OJS version. This meant that even if the Septentrio team put in extra work hours to take care of the translation files in the following week, authors, editors and reviewers of Septentrio journals would be met with unfamiliar features and possible upgrade hiccups.

In Norway they say: “Det er ingen skam å snu” – there is no shame in turning back. So we terminated this trip and turned back to the PDF-only format. Next time we are better prepared.

Lessons learned

Despite the disappointing failure, we now have a good understanding of open source options for XML-based publishing and the relative strengths and limitations of the different available solutions. We have developed a workflow that, although imperfect and still requiring some manual corrections, could be used to develop an implementation tailor-made for editors who want to incorporate XML into their production processes. A more thorough and extensive future project could include the following:

- Find out if any other institutions have undertaken similar projects and are willing to share their experiences and insights with us.
- Investigate and compile a list of all OJS-compatible open source tools for converting Septentrio scholarly documents to XML.
- Review, test and evaluate the available tools on latest version of OJS before upgrading.
- Edit XML files and render XML into PDF and HTML.
- Validate XML files resulting from different tools.
- Analyze existing documents in Septentrio (published journal articles) to identify potential problems.
- Test tools with DOCX, Google Doc, OpenOffice, and LibreOffice sample files. LaTeX files should also be tested.
- Try to combine the tools into a workflow and take note of points that need attention.
- Identify areas of improvement for the tools selected.
- Produce an initial guide and template for authors and editors.