

## **Notat - Håndbok for bruk av metadata i norske forskningsarkiv**

Styringsgruppa for NORA og NORA-lederen oppnevnte i sin tid undertegnede som arbeidsgruppe for å utarbeide et utkast til en felles håndbok med anbefalinger for registrering av metadata i åpne norske forskningsarkiver. Arbeidsgruppa legger med dette frem sitt forslag til en slik håndbok. Vårt forslag sendes over til Cristin for videre behandling.

I utkastet til håndbok gir vi våre anbefalinger om bruk av kvalifiserte Dublin Core-elementer og registreringspraksis av disse. Vi ønsker i forbindelse med oversendelsen å utdype våre vurderinger på noen punkter.

### **Registrering av tittel og undertittel:**

Vi har vurdert to måter å registrere tittel og undertittel på:

1. Registrere tittel slik det anbefales i AACR, dvs. at hoved- og undertittel skilles med : (kolon)
2. Tittel registreres som beskrevet i ressursen. Hvis det ikke finnes noe skille tegn mellom over- og undertittel kan punktum, bindestrek registreres (valgfritt).

Arbeidsgruppa har valgt alternativ 2 av hensyn til brukervennligheten. Det er slik tittelen er registrert på ressursen og slik forfatter selv vil forvente å få opp tittel ved søk. Det vil være uheldig å registrere tittel på annen måte i metadata enn slik den fremkommer i fullteksten som følger med innførselen. Mange artikler vil overføres fra Cristin, som igjen får artikkeldata fra utgiver. Det vil være naturlig å benytte de samme dataene i arkivene som i Cristin. Endringer som skal gjøres i arkivene, som alternativ 1 vil innebære, vil dessuten innebære noe merarbeid.

### **Opphavsmann og bidragsyter:**

Vi anbefaler at creator brukes for den som skaper et verk (juridisk opphavsperson), mens contributor brukes for de som bidrar til et verk, men som ikke har opphavsrett til det. Vi har valgt å kvalifisere dc.creator.author for å få et felt som kan brukes spesielt til forfatter (og få etikett forfatter). Andre opphavspersoner kan da kvalifiseres i egne creator-felt.

Vi har vært usikre på i hvilken grad vi skulle anbefale en mer uttømmende liste med kvalifikatorer for elementene creator og contributor. Vi har valgt å foreslå en ikke-uttømmende liste som kun omfatter de kvalifikatorene som er mest brukt og i stedet valgt å benytte dc.creator.other og dc.contributor.other som "sekkeposter" for det som ikke dekkes av kvalifikatorene. Arbeidsgruppa mener det er bedre om institusjonene heller legger til egne kvalifikatorer for dc.creator og dc.contributor dersom de har lokale behov for det, enn å operere med en uttømmende liste i håndboka.

### **Ressurstype:**

Vi har valgt å anbefale bruk av dc.type.resource for å beskrive ressurstype. Dette gjør det mulig å beskrive en ressurs bedre på filnivå. DCMI har en anbefaling av et vokabular for elementet type <http://dublincore.org/documents/dcmi-type-vocabulary/>, som kan brukes for å beskrive ressurstyper. Disse er i dag i bruk i ODA og TEORA.

Dc.resource er et tillegg til bruk av dc.format. Dc.format genereres automatisk av DSpace på filnivå (bitstream). NTNUs Diva benytter tilsvarende. dc.format anbefales også av DRIVER. dc.format har fire hovedtyper med underkategorier: text, application, image, audio og video. dc.format sier imidlertid ikke

noe om den fysiske, digitale manifestasjon på filnivå. For eksempel vil en PDF-fil kategoriseres som applikasjon og ikke som tekst.

#### **Populærvitenskapelig sammendrag:**

Vi anbefaler en ny kvalifikator (dc.description.popularized) for å kunne legge inn popularisert sammendrag i tillegg til et vitenskapelig sammendrag. Det er primært for doktorgradsavhandlinger at det er meldt et behov for denne kvalifikatoren, men den kan selvsagt også brukes for andre dokumenttyper. Et slikt felt bidrar til å styrke formidlingen av doktoravhandlingene og vitenskapen.

#### **Fagfellevurdering:**

Vi foreslår å opprette et eget felt for fagfellevurdering - dc.type.peerreviewed. Tanken er at Peer reviewed skal kunne trekkes ut som et eget ja/nei avkrysningsfelt i arbeidsflyten. Dette betyr samtidig at "Fagfellevurdert / Peer reviewed" fjernes i listen over dokumenttyper i NORA-standard, jevnfør [http://www.ub.uit.no/wiki/openaccess/images/1/15/Dokumenttyper\\_NORA\\_09.pdf](http://www.ub.uit.no/wiki/openaccess/images/1/15/Dokumenttyper_NORA_09.pdf).

Formatted: Font: 10 pt

Formatted: Font: 10 pt

#### **Bibliografisk sitering:**

Det er kanskje det felte arbeidsgruppa har vært aller mest i tvil om. DCMI anbefaler å bruke dc.identifier.bibliographiccitation mens DRIVER anbefaler bruk av feltene dc.source (for digitalisert materiale)<sup>1</sup> og dc.relation (for digitalt originalmateriale).

Formatted: Font: 10 pt

Formatted: Font: 10 pt

En stor utfordring med Dublin Core er at standarden ikke har støtte for f.eks. årgang eller sidetall for tidsskriftsartikler, og i stedet legger all siteringsinformasjon i ett enkelt felt, nemlig dc.identifier.bibliographiccitation. På akkurat dette punktet foreslår vi å avvike fra Dublin Core standarden fordi det er viktig å ha muligheten til gjenbruk av data. Uten en slik kvalifisering blir det vanskeligere å eksportere referanser til referansehåndteringsverktøy som EndNote, Reference Manager og Zotero. Her har imidlertid Helsebiblioteket's Research Archive (HeRA) brukt dc.identifier.bibliographiccitation til å eksportere til EndNote, RefMan og RefWorks. En slik løsning krever imidlertid nøyaktig registrering i referansestrengen som blir lagt inn for at den skal fungere.

Vi foreslår derfor at identifier kvalifiseres med egne felter for tidsskriftartikler (tidsskriftets navn, volum, hefte, sidetall for hvor artikkelen starter og sidetall for hvor artikkelen slutter).

For andre dokumenttyper som bokkapitler, konferanseinnlegg m.m. foreslår arbeidsgruppa at feltet dc.relation.isPartOf brukes. Her registreres navn på serie ressursen inngår i. For bokkapitler registreres her navn på boken. I tillegg kan de kvalifiserte identifierfeltene brukes for å registrere volum, nummer og sidetall.

Vi anbefaler bruk av dc.identifier og dc.relation.isPartOf som nevnt ovenfor, men anbefaler samtidig at dette er noe som testes nærmere av ett eller flere arkiv før det tas i bruk.

#### **OpenAIRE-anbefalingene:**

Det bør tas høyde for anbefalingene fra OpenAIRE i arkivene, og arbeidsgruppa har derfor innarbeidet OpenAIRE-kravene i utkastet til håndboken.

---

<sup>1</sup>Anbefalingen av dc.source er uklar hos DRIVER siden de samtidig henviser til DCMI sin anbefaling om å bruke dc.identifier.bibliographicCitation. Se DRIVER Guidelines s. 75.

Dette betyr at vi anbefaler å bruke dc.date.embargoEndDate og at denne kvalifikatoren erstatter de embargofeltene som brukes av Dspace i dag (dc.embargo.liftdate og dc.embargo.terms).

Vi anbefaler også at feltet dc.relation.projectID tas i bruk. Dette feltet er obligatorisk for forskning som er omfattet av OpenAIRE-kravene, dvs utvalgte deler av EUs 7. rammeprogram.

Mange av dataene som skal brukes av OpenAIRE vil formodentlig komme fra Cristin. Nøyaktig hvordan integrasjonen mellom arkivene, Cristin og OpenAIRE skal foregå må avklares.

#### **Registrering av doktorgradsavhandlinger med artikler:**

Denne problemstillinga er ikke beskrevet i DRIVER Guidelines og vi står derfor ganske fritt til å etablere egen praksis. Hva vil være den beste løsningen for dette?

Arbeidsgruppa foreslår at feltet dc.relation.hasPart brukes til å registrere artiklene som inngår i artikkelbaserte doktoravhandlinger. Feltet brukes både for artikler som kan legges i arkivet (lenke til URI i arkivet), og for artikler som ikke kan legges i arkivet (lenke til URI hos utgiver, dersom mulig). Dette er slik det blant annet gjøres i BORA i dag. Problemet med valg av denne løsningen er at html-koder da legges inn i et felt som i utgangspunktet ikke er tilpasset dette formålet.

En alternativ løsning kunne være å registrere alle artikler som inngår i en doktoravhandling som egne innførsler i arkivet, selv når det ikke er mulig å legge ut artikkelen som fulltekst. I feltet dc.relation.hasPart kunne man så lenke til arkivposten for hver enkelt artikkel (og også lenke den andre veien via dc.relation.isPartOf). Dette alternativet brukes av NTNUs DIVA i dag. Et minus med denne løsningen er at arkivet dermed ikke vil være 100% Open Access.

Med vennlig hilsen

Ingrid Cutler, UB Bergen  
Arild Skalmeraas, Høgskolen i Telemark, biblioteket  
Kristian Seterås, Havforskningsinstituttet, biblioteket  
Leif Longva, UB Tromsø