

Danske biblioteker i WorldCat

30. august 2010

Adgang til og genbrug af katalogposter fra WorldCat



Indledning

I 2008 blev der med Statsbiblioteket, DBC og Styrelsen for Bibliotek og Medier som hovedaktører gennemført et projekt, der skulle forberede et nærmere samarbejde mellem danske biblioteker og den internationale biblioteksbase WorldCat. Udredningsarbejdet mandede ud i, at der i december 2008 blev underskrevet en aftale, der dels gav danske biblioteker adgang til at bruge WorldCat som kilde til genbrugskatalogisering, dels indebar, at den samlede bestand på danske offentlige biblioteker skulle uploades til WorldCat.

Aftalen mellem OCLC og DBC betyder bl.a. at alle offentlige danske biblioteker uden ekstraomkostninger frit og ubegrænset kan søge i WorldCat samt downloade og genbruge poster fra WorldCat til eget brug.

For at sikre, at danske biblioteker får den maksimale nytte af de nye muligheder blev der året efter igangsat et nyt DEFF projekt *Danske biblioteker i WorldCat – implementering 2009*. Heri indgår bl.a., at der på baggrund af tests og opsamlede erfaringer skal udarbejdes et sæt guidelines og/eller arbejdsbeskrivelser, som kan gøre det lettere for både folke- og forskningsbiblioteker at genbruge katalogposter fra WorldCat.

I flere biblioteker er anvendelsen af WorldCat som datakilde gnidningsløst blevet en del af de daglige katalogiseringsrutiner, og der er tydeligvis ikke behov for at udarbejde en egentlig manual på området. Alligevel vil man flere steder kunne have glæde af en minivejledning, der redegør for nogle af de særlige karakteristika ved WorldCat-posterne og udpeger områder, hvor der kan være grunde til at rette downloadede poster og tilnærme dem dansk formateringspraksis.

Nærværende vejledning bygger på bidrag fra Henning Midtgaard Hanssen, Statsbiblioteker og Bodil Dalgaard-Møller, DBC. Vejledningen har været til høring hos projektets samarbejdspartnere og er efterfølgende redigeret af Per Mogens Petersen, DBC og Leif Andresen, Styrelsen for Bibliotek og Medier.

Styregruppen for Danske biblioteker i WorldCat – implementering 2009
30. august 2010

WorldCat – basens indhold

WorldCat drives af den USA-baserede organisation OCLC (Online Computer Library Center) og er en fælleskatalog for biblioteker i hele verden - flest dog i USA. I WorldCat er der registreret bøger, periodika, musikoptagelser, noder, websites, film og video, computer programmer og manuskripter - basen er i juni 2010 på 193 mio. poster. Over halvdelen af alle poster er nu på andre sprog end engelsk.

Ultimo 2009 eksporteredes ca. 13 mio poster fra DanBib/bibliotek.dk, heraf blev 7.8 mio poster indlagt som "nye" poster i WorldCat.

Der henvises til OCLC for yderligere oplysninger om indholdet i basen. Der er ikke ét sted hvor alle oplysninger står, men på <http://www.oclc.org/worldcat/default.htm> er der nogle informationer.

Genbrug af poster fra WorldCat

DanBib-licensen giver alle offentlige danske biblioteker adgang til genbrug af poster fra WorldCat via Netpunkt og Zpunkt. Benytter man adgangen via Zpunkt, skal man være opmærksom på, at funktionaliteten vil være påvirket af bibliotekets genbrugs- og katalogiseringsklient.

DanBib stiller de sædvanlige muligheder for download af poster til rådighed – se nærmere på: <http://www.danbib.dk/index.php?doc=download>

Posterne kan konverteres "on-the-fly" fra MARC21 til danMARC2. I Netpunkt kan man alternativt vælge at få vist posterne i MARC21.

Uanset valg af format, katalogiseringsklient og konverteringsprogram skal man være opmærksom på, at en post, som downloades fra WorldCat, **skal** indeholde et OCLC-ID, og at dette ID-nummer ved den efterfølgende overførsel til DanBib **skal** stå i felt 035 *a med en særlig indledende tekst: (OCoLC).

Eksempel: **035 00 *a(OCoLC)173198721**, jf. danMARC2-formatet.

Uanset om biblioteket anvender Netpunkt eller Zpunkt skal dette respekteres, idet den økonomiske del af aftalen mellem OCLC og DBC er baseret på antallet af poster, som danske biblioteker genbruger fra WorldCat. Genbrugets omfang beregnes af DBC ved regelmæssig optælling af antallet af poster i DanBib med OCLC-ID i felt 035. Det er derfor vigtigt, at dette felt findes i alle poster som er genbrugt fra WorldCat.

Når biblioteket genbruger poster fra WorldCat og henter dem i danMARC2 via DBCs konvertering, oprettes **automatisk** korrekt formaterede 035-felter.

Søgning i WorldCat via Netpunkt

Når man søger i WorldCat via Netpunkt, søger man i OCLC's database, men med Netpunkts brugergrænseflade.

WorldCat er i forhold til DanBib kendetegnet ved at posterne er formateret i MARC21 og med et andet tegnsæt. DBC kan ikke ændre i hverken søgemuligheder eller data, men har tilpasset Netpunkts brugergrænseflade med navne på og definition af

søgekoder, visformater incl. ledetekster, tegnkonvertering etc. så det mest muligt svarer til søgning i DanBib via Netpunkt.

Fra Netpunkt kan man vælge at **søge direkte i WorldCat via vælg database**, eller man kan fx starte sin søgning i DanBib og derefter udføre en **videresøgning** i andre databaser, herunder WorldCat.

Fortrolighed med WorldCat opnås bedst ved at bruge den direkte vej via **vælg database**, idet søgemuligheder mv. er beskrevet på forsiden. Disse tekster vises ikke ved videresøgning fra andre databaser.

Posterne fra WorldCat konverteres "on the fly" fra MARC21 til danMARC2 - så godt som dette nu lader sig gøre, dvs. søgning foretages i MARC21-formaterede poster, medens resultatet vises i danMARC2-formaterede poster.

Se evt. også vejledning i brug af WorldCat via OCLC's egen brugergrænseflade, FirstSearch:

http://www.oclc.org/support/documentation/firstsearch/z3950/z3950_databases/specs/worldcat.htm.

Genbrug af poster via Netpunkt

Fra Netpunkt kan poster downloades på 2 forskellige måder:

- 'accept'-metoden: med 'accept' overføres posterne løbende til en fil på DBC's Posthus. Herfra overføres de til det lokale bibliotekssystem, enten via Netpunkt's Posthus-adgang eller via lokalt ftp-program
- 'hent'-metoden: med 'hent'-metoden overføres posterne til det lokale bibliotekssystem ved hjælp af e-mail

Valg af 'accept' eller 'hent' afhænger af det lokale bibliotekssystem. Husk at aktivere link til 'accept' eller 'hent' via fanebladet **Opsætning** på netpunkt.dk.

Adgang til WorldCat via bibliotekets egen klient

Mange biblioteker har valgt at benytte WorldCat gennem egen brugergrænseflade / klient.

Søgning og visning

Det anbefales at søge enten på ISBN/ISSN eller på kombinationer af enkeltord (fx fritekst eller enkeltordssøgeregistre). Skan-søgning er ikke mulig fra alle katalogiseringsklienter og giver i bedste fald et usikkert resultat – især hvis man foretager samsøgning med andre baser.

Hvis katalogiseringsklienten tillader CCL-søgninger mod WorldCat, kan man søge på materialetype med søgekoden *ma*. Som materialetypekoder bruges de amerikanske, der findes i OCLC-dokumentationen afsnittet Material type:

http://www.oclc.org/support/documentation/firstsearch/z3950/z3950_databases/specs/worldcat.htm, fx: ti=rinaldo og ma=rec.

Selv om tegnene æ/ä og ø/ö vises korrekt i posterne, skal de ofte søges på anden vis – prøv *a* for æ/ä og *o* for ø/ö. Fx søges Eckernförde som Eckernforde.

Der er også hjælp at hente i DanBib dokumentationen, fx

http://www.danbib.dk/index.php?doc=zpunkt_mod_eksterne#soeg

Posterne som hentes til genbrug fra WorldCat er uden lokaliseringsoplysninger. Den rækkefølge, hvori posterne vises, er imidlertid ofte bestemt af antallet af lokaliseringer.

Opsætning af klienten

Klienten skal opsættes til WorldCat via Zpunkt, i praksis kaldes **base WorldCat** på z3950.dbc.dk port 210. Posterne kan hentes i 3 "present"-formater: danMARC2, MARC21 og SUTRS.

Man kan søge med de søgekoder og attributkombinationer som WorldCat Z39.50-targetet stiller til rådighed. Adgang via Zpunkt forudsætter derfor en specifik opsætning af Z39.50-klienten.

Generelt henvises til dokumentationen på <http://www.danbib.dk>. Om download via Zpunkt ved hjælp af en Z39.50-klient, se specielt <http://www.danbib.dk/index.php?doc=zpunkt>. Som hjælp til opsætning henvises til DBC's opsætning til netpunkt.dk: <http://www.danbib.dk/docs/worldcatsearch.htm> og <http://www.danbib.dk/docs/worldcatscan.htm>.

OCLC's Z39.50-dokumentation er omfattende, men start evt. her: http://www.oclc.org/us/en/support/documentation/z3950/config_guide/. De registre (use-attributter) Z39.50-adgangen stiller til rådighed kan ses på: <http://www.oclc.org/support/documentation/z3950/searchtips/>

Datamodel

OCLC opfordrer de institutioner, som leverer data til WorldCat, til så vidt muligt at genbruge poster, som allerede er skabt af andre, men beskriver dog på hjemmesiden en række situationer, hvor det anses for rimeligt at oprette en ny post: <http://www.oclc.org/us/en/bibformats/en/input/default.shtm>

I praksis bliver poster fra Danmark aldrig matchet med en oprindelig OCLC-post; matchning med den oprindelige OCLC-post gælder kun biblioteker som katalogiserer direkte i WorldCat.

Felt 040 indeholder (i delfelt a) OCLC's kode for det bibliotek, som har oprettet posten i WorldCat. Poster fra DanBib indeholder enten DKDLA eller DED; DKDLA er poster fra DanBib fra den generelle leverance, medens DED er poster fra tidligere specielle leverance af DPB-poster.

OCLCs matchning af poster afspejles i felt 029. Antallet af 029-felter i en given post svarer til antallet af poster som er blevet matchet af OCLC; indholdet af delfelt a er også i dette tilfælde OCLC's kode for det bibliotek som har uploadet posten til WorldCat, dvs. at poster fra Danmark enten har DKDLA eller DED i 029 delfelt a. Første indikator = 0 markerer den oprindelige post, dvs. er identisk med den kilde som også er angivet i 040. Øvrige 029-felter markerer så de poster, der er blevet matchet med den.

Resultatet er, at det ikke er usædvanligt at finde 5-6 (undertiden endnu flere) forskellige poster for det samme dokument.

Både 040 og 029-felterne slettes ved DBC's konvertering fra MARC21 til danMARC2, men kan altså ses i MARC21.

Katalogiseringsniveau

De fleste poster i WorldCat ligger på, hvad man kunne kalde et nationalbibliografisk niveau, fordi Library of Congress og en række nationalbiblioteker er blandt de største dataleverandører.

Nu og da kan man dog møde poster, der ser ud som foreløbige poster eller accessionsposter. De mangler typisk angivelse af sidetal eller tilsvarende fysisk beskrivelse og anfører alle forfatternavne i ubearbejdet form i felt 720.

Koden for katalogiseringsniveau i felt 008 *v er generelt upålidelig. De netop omtalte foreløbige poster har som regel koden 3 (abbreviated level), mens fuldt udbyggede poster fra Library of Congress sjældent når højere end niveau 4 (core level).

Emneord

De fleste WorldCat-poster indeholder (efter konvertering til danMARC2) emneord i et eller flere af disse felter:

- 600 Personnavn som emneord
- 610 Korporationsnavn som emneord
- 630 Kontrolleret emneord
- 650 Library of Congress emneord og analoge emneord
- 651 Library of Congress emneord og analoge emneord indledt med geografisk navn
- 660 MeSH¹ emneord
- 662 AGROVOC²
- 670 NAL emneord³

Hovedparten af disse emneord er på engelsk, men der er et voksende antal emneord på andre sprog: fransk, tysk, hollandsk, spansk m.fl.

Emneordsfelter, der har deres oprindelse i de nationale amerikanske biblioteksinstitutioner, er ikke mærkede på nogen særlig måde, mens andre sædvanligvis indeholder et delfelt *2 med en kode, der angiver kilde eller emneordssystem. Hvis man ønsker at holde fx emneord på hollandsk ude af sin katalog, kan man frasortere dem på grundlag af de omtalte koder. Koderne⁴ kan ses på: <http://www.loc.gov/standards/sourcelist/subject.html>

Indholdsnoter

1 MeSH: Medical subject headings

2 AGROVOC: Multilingual agricultural thesaurus

3 NAL: National Agricultural Library (USA)

4 De talkoder fra 1 til 5, som i danMARC2 er godkendt til brug i emneordsfelternes *2, benyttes ikke i DBC's konvertering.

WorldCat-posterne er ofte forsynet med fyldige indholdsnoter i felt 530 (sjældnere i andre notefelter). I nogle tilfælde er oplysningerne formateret i de specifikke delfelter *i *t *e osv., i andre blot i ét sumdelfelt *a. For musikoptagelser betyder det, at det ikke generelt er muligt at udtrække data for enkelte skæringer eller tracks.

Mange vil sandsynligvis se det som en hindring for genbrug af musikposter, at danMARC2's brug af analysefelter (770, 780,795) kombineret med en indledende tekst til en maskingenereret indholdsanalyse i felt 531 er fremmed for MARC21 og dermed også for WorldCat-posterne. Desværre er det et problem, man ikke kan konvertere sig ud af!

Flerbindsværker

De fleste WorldCat-poster for flerbindsværker er formateret efter den såkaldte ét-niveau-metode, hvor oplysningerne om de enkelte bind er placeret i felt 530. Inden for feltet kan oplysningerne være anbragt i ét delfelt *a og adskilt ved hjælp af typografiske skilletegn, eller de kan være formateret i flere delfelter (typisk titel i *t, ophavsangivelse i *e og bindnummer eller anden supplerende tekst i *b). Undertiden møder man også poster inddateret efter den såkaldte bibliografi-metode. Her findes en post for hvert bind eller gruppe af bind, der indeholder både fællesoplysninger (normalt i felt 245 *a) og nummer og titel på det enkelte bind (i felt 245 *n og *o).

De metoder, der foretrækkes her i Danmark – to-niveau-metoden med felt 248 og fler-niveau-metoden med hovedpost, evt. sektionspost og bindpost – er ukendte i WorldCat.

Lokale data

Til poster importeret fra WorldCat kan man naturligvis uden indskrænkninger føje lokale felter inden for områderne 087-099 og 560-599 samt – for periodica – beholdningsoplysninger i felt 980 (og 985).

Lokale emneord kan fortsat inddateres i felterne 600, 610, 630, 631, 633, 634 og 645.

Enkelte felter

Nedenfor gennemgås specifikt for enkelte felter forskellige forhold, man bør være opmærksom på i WorldCat-posterne.

Felt	Bemærkninger
009	Kontroller, at koden for specifik materialebetegnelse svarer til det forhåndenværende materiale
035	Feltet må aldrig slettes (se side 2)
041	Kontroller, at relevante sprogkoder er inddateret
250	Kontroller, at feltet indeholder, hvad vi normalt forstår ved en udgavebetegnelse
538	Kontroller, at numre for forskellige udgaver eller versioner ikke er blandet sammen i samme post
600, 610 650, 651	Delfelt *2 indeholder MARC21's koder for anvendt emneordssystem. Det er bogstavkoder, som er forskellige fra de godkendte talkoder i danMARC2
660	Kontroller, at delfelter, der er tænkt som led i en streng, ikke er fordelt på hvert sit felt

- 700, 710 Kontroller navneformer i forhold til danske katalogiseringsregler.
Parentetiske tilføjelser – typisk (*Musical group*) - slettes eller oversættes til dansk
- 720 Kontroller, at navne er forsynet med funktionskoder, hvis de ikke indgår i den bibliografiske beskrivelse (primært som 245 *f)

Administrative MARC21-felter

Visse MARC21-felter illustrerer datamodellen i WorldCat, herunder match og ajourføring af poster. Nedenstående oversigt er ikke eksakt, men et forsøg på at beskrive det mest almindeligt forekommende tilfælde.

OCLC sammenlægger typisk **ikke** poster fra forskellige kilder. Amerikanske biblioteker katalogiserer normalt direkte til WorldCat, i hvert fald optræder de fleste lokaliseringer til amerikanske biblioteker i samme post. Ved siden af den amerikanske post er der så (afhængigt af den pågældende titel) typisk fx en post fra Frankrig, en fra Tyskland og en fra Danmark; da DBC har eksporteret samtlige poster fra bibliotek.dk forekommer der i visse tilfælde endog mere end én post med danske lokaliseringer – det er naturligvis en konsekvens af vores datamodel, hvor alle forskningsbiblioteker hver har deres egne poster i DanBib med dertil hørende risiko for forskelle som forhindrer match hos OCLC.

I poster, som afsendes fra DanBib/bibliotek.dk, opretter/indsætter DBC indeholder følgende "administrative" MARC21-felter:

Poststatus	n, c eller d	ny post, rettet post eller slettepost. Indsættes i postetiketten (record label); poster slettemarkeret er poster slettet af et dansk ejerbibliotek og det medfører sletning af den tilhørende beholdningspost i WorldCat, i praksis sletning af et 029-felt, jf. nedenfor.
DFA-ID	001 xxxxxxxxxx	Internt ID-nummer i bibliotek.dk (DFA-ID) indsættes i felt 001
OCLC-ID	035--*a(OCOLC)yyyyyyyy	OCLC-ID hentes fra DBC's krydsreferenceliste WCX og indsættes i felt 035. Feltet medsendes ikke ved nye poster.
Katalogiseringskilde	040--*adan	Indsættes i alle poster
ISIL-kode for lokalisering	049--*aDK-190101 eller 049--*aDK-190101*aDK-810010	Altid lokalisering til bibliotek.dk (DK-190101) samt evt. ét af RS-bibliotekerne, i eksemplet Det kongelige Bibliotek

Når OCLC har behandlet posterne fra Danmark, fjernes eller ændres ovenstående felter, således at WorldCat indeholder bl.a. følgende "administrative" MARC21-felter:

OCLC-ID	001 yyyyyyyy	ID-nummeret er genbrugt eller nyt afhængigt af poststatus og forekomsten af 035, jf. ovenfor
---------	--------------	--

Katalogiseringskilde	040--*aDKDLA*bdan*cDKDLA eller 040--*aDED*bdan*cDED	DKDLA er OCLC's symbol for bibliotek.dk. DED er OCLC's symbol for DPB (fra ældre leverancer)
Eksterne numre – set fra OCLC	0290-*axxxxxxxxx 0291-*azzzzzzzzz 0291-*awwwwww etc.	029 indeholder (i poster fra DK) interne ID-numre fra bibliotek.dk (DFA-ID) Første indikator = 0: primær post, dvs. den der er lagt ind i WorldCat. Første indikator = 1: sekundære poster; kun ID-nummeret er lagt ind i WorldCat-posten

Til ét OCLC-ID vil ofte høre mere end ét DFA-ID, jf. antallet af 029 i OCLC-posten.