

## A. WU IN A. MITCHELL: MASS MANAGEMENT OF E-BOOK CATALOG RECORDS

### UPRAVLJANJE S PAKETI BIBLIOGRAFSKIH ZAPISOV ZA E-KNJIGE

V prispevku predstavljamo članek avtorjev Annie Wu in Anne Mitchell z naslovom *Mass Management of E-Book Catalog Records: Approaches, Challenges, and Solutions*, ki je bil objavljen leta 2010 v publikaciji *Library Resources & Technical Services*, 54(3), 164–174.<sup>1</sup>

Uporaba elektronskih knjig (e-knjige) se v zadnjih letih hitro povečuje (Amazon je pred kratkim objavil novico, da je prodaja e-knjig preseгла prodajo trdo vezanih izdaj<sup>2</sup>), kar postavlja knjižničarje in katalogizatorje pred zahtevno nalogo – omogočiti hiter, ažuren in učinkovit dostop do najnovjših vsebin ter hkrati zavarovati pravice uporabnika in avtorjev. Ob tem pa je seveda treba spoštovati licenčne pogoje v pogodbah z založniki, ki se trenutno zastrujejo do nesmisla,<sup>3</sup> kar onemogoča prost razvoj vseh možnosti, ki jih knjigi daje digitalizacija,<sup>4</sup> in žal potrjuje nekatere zelo tehtne pomisleke uporabnikov.<sup>5</sup>

Annie Wu, koordinatorica katalogizacije, in Anne M. Mitchell, vodja katalogizacije in metapodatkovnih servisov v univerzitetni knjižnici University of Houston Libraries (UHL),<sup>6</sup> v članku *Mass Management of E-Book Catalog Records: Approaches, Challenges, and Solutions* predstavljata postopke, povezane s katalogizacijo e-knjig. Na osnovi pregledane literature uvodoma ugotavljata, da obstaja širok konsenz o nujnosti vključevanja bibliografskih zapisov (bibl. zapisov) za e-knjige v lokalne kataloge. Veliko knjižnic se pri katalogizaciji velikih zbirk (nekaj desetisoč naslovov) zanaša na bibliografske zapise, ki jih ob nakupu zagotavljajo dobavitelji. In kot se med dobavitelji razlikujejo vsebine in formati njihovih zbirk, način dostopa in pravice uporabnikov, tako se razlikuje tudi vsebina posameznih polj v bibliografskih zapisih, ki jih ob tem zagotavljajo. Kljub razširjenemu vključevanju bibliografskih zapisov v lokalne baze pa so se priporočila za katalogizacijo e-knjig do sedaj osredotočala predvsem na vsebino posameznih polj in zapisov, na logistiko upravljanja zelo obsežnih paketov teh zapisov, ki imajo veliko fluktuacijo, pa se je pozabljalo.

Poleti 2009 je delovna skupina v okviru mednarodnega projekta "Program for Cooperative Cataloging" (PCC)<sup>7</sup> izdala priporočila glede bibliografskih zapisov za e-

knjige, neodvisnih od dobavitelja "Provider-Neutral E-Monograph MARC Record Guide: Guidelines and Metadata Application Profile (MAP).<sup>8</sup> Po mnenju avtoric članka pomenijo priporočila velik korak naprej v standardizaciji katalogizacijske prakse za e-knjige, vendar pa ugotavljata, da je (ponovno) premalo pozornosti namenjene integraciji obsežnih zbirk bibliografskih zapisov različnih dobaviteljev.

Glavni izzivi, s katerimi se pri postopku vključevanja bibliografskih zapisov soočajo knjižnice UHL, so:

- paketno urejanje bibliografskih zapisov,
- združevanje podvojenih bibliografskih zapisov,
- načrtovanje in sledenje posodobitvam posameznih zbirk,
- pridobivanje in vzdrževanje znanja in veščin katalogizatorjev, ki opravljajo te naloge.

Avtorici v posameznih poglavjih predstavita trenutni potek katalogizacije e-knjig v UHL – odločitve pri lokalni "paketni" katalogizaciji, uporaba odprtokodnega orodja "MarcEdit"<sup>9</sup> za upravljanje z bibliografskimi zapisi v formatu MARC, uporaba programskega segmenta "MARC service for e-books" SerialsSolutions<sup>10</sup> in koordinacija procesa paketnega upravljanja z zapisi.

Razvoj trga e-virov je prinesel dostopnost širokega nabora monografskih e-vsebin – referenčnih del, akademskega in tehničnega gradiva ter literarnih in primarnih virov. Podobno raznolik je način nabave teh vsebin – nabava posameznih naslovov, paketi založnikov in drugih dobaviteljev ter predlogi uporabnikov. Avtorici ugotavljata, da največja težava ob tem ni sama katalogizacija, temveč učinkovito upravljanje celotnega procesa obdelave. Zbirke e-knjig se namreč zelo hitro spreminjajo, bibliografski podatki, ki jih spremljajo, pa prihajajo iz zelo različnih virov in so oblikovani na osnovi različnih standardov. Dodajmo, da ob tem postaneta ustrezno vzdrževanje in kvaliteta bibliografskih zapisov v lokalni bazi še pomembnejša.

Nabava e-knjig lahko poteka na osnovi različnih modelov, posamezno ali v paketu, vsak model pa po svoje

vpliva na katalogizacijo in upravljanje bibliografskih zapisov. Proces obdelave pri posameznem naslovu je podoben obdelavi tiskanih knjig. Kot izziv pri katalogizaciji je izpostavljeno sledenje posameznim korakom v procesu obdelave gradiva in kdo je zanj odgovoren na določeni stopnji.

Akadske knjižnice večinoma nabavljajo e-knjige v paketih. Najlažje je upravljanje s statičnimi paketi, kot so npr. različne zbirke netLibrary, saj je zapise treba naložiti le enkrat, nato pa dopolnitve niso več potrebne. Relativno enostavno je tudi upravljanje paketov leposlovnih del in primarnih virov, saj so dopolnitve redke in se dodajajo k obstoječemu paketu – opuščanje naslovov je pri teh vrstah paketov zelo redko. Največji izziv za upravljanje pa predstavljajo paketi, katerih vsebina se stalno spreminja, kot je npr. zbirka Safari Tech Books, ki vsebuje dela s tehnično vsebino, objavljena v zadnjih treh letih. Dopolnitve in opuščanje naslovov se dogajajo mesečno in proces upravljanja z zapisi hitro postane zelo težaven, še posebej, če ima knjižnica naročenih več takšnih zbirk. Za učinkovito upravljanje z zapisi avtorici poudarjata najpomembnejšo vzpostavitev sekundarne infrastrukture – sistema za upravljanje z elektronskimi viri (ERM), tabelami, vikiji ali njihovo kombinacijo.

Enostavnost bibliografskega upravljanja je v veliki meri odvisna od velikosti, narave in spremenljivosti paketa. Avtorici navajata več različnih tipov vsebin:

- *Tehnične in profesionalne knjige* (npr. ENGnetBASE, Safari Books Online, Digital Engineering Library) hitro zastarijo, starejši naslovi so periodično nadomeščeni s podobnimi ali novejšimi izdajami. Zaradi tehničnih vsebin so informacije o datumu in izdaji zelo pomembne za uporabnike.
- *Referenčne knjige* (npr. Credo Reference, Oxford Reference Online, Sage eReference) so dostopne posamezno ali v manjših paketih. Tudi referenčni paketi se pogosto dopolnjujejo z novimi izdajami. Nekatera spletna referenčna dela se stalno dopolnjujejo, podobno kot integrirni viri.
- *Paketi literarnih del in primarnih virov* (npr. baze podatkov Chadwyck-Healy in Alexander Street Press) so relativno statični, zelo verjetno pa vsebujejo neknjižno monografsko gradivo, ki zahteva drugačno obravnavo (kratke fantastične zgodbe, poezija, drame in primarni viri kot so pisma, intervjuji in dnevniki).
- *Paketi z vsebino več založnikov* (ebrary, netLibrary) so običajno zelo obsežni, zajemajo veliko področij in so lahko statični ali dinamični. Največji izziv pri upravljanju s takšnimi paketi je njihova velikost, saj globalne spremembe zahtevajo veliko sistemskih zmogljivosti.

Takšna raznolikost vsebin in obsega kar kliče po standardizaciji in stanovitnem načinu upravljanja.

Priporočila PCC glede bibliografskih zapisov za e-knjige uvajajo model enega originalnega (angl. *master record*) zapisa, ki (namesto ločenih zapisov za verzije posameznih dobaviteljev) obsega vse enakovredne pojavne oblike nekega naslova e-monografije.

Takšen pristop po mnenju avtoric prinaša dve pomembni prednosti. V okolju vzajemne katalogizacije ustavlja širjenje najrazličnejših zapisov za isto vsebino. Zmanjšanje števila in povečanje konsistentnosti zapisov za e-knjige pa katalogizatorjem močno olajša iskanje in identifikacijo ustreznih zapisov. Delo, porabljeno za iskanje in pregledovanje številnih podobnih zapisov ali vnos novih zapisov, ki so zelo podobni že obstoječim, se tako lahko nameni dopolnjevanju originalnega zapisa s predmetnimi oznakami, opisom vsebine in normativno kontrolo.

Nova priporočila več ne delajo razlik med reprodukcijo in izvirno digitalno monografijo in prinašajo jasna navodila za uporabo posameznih polj iz bibliografskih zapisov v formatu MARC,<sup>10</sup> ki so se pred tem uporabljala nekonsistentno (polja 534, 773, 776). Datumi in informacije o založniku izhajajo iz izvirne monografije (tiskane ali elektronske), kar je v skladu z izbiro uporabnikov, ki v območju založništva želijo informacije o izvirnem založniku in času izdaje.

Večina polj za opombe se opušča, z izjemo nabora opomb o elektronski obliki (monografije), ki izhaja iz Digital Library Federation Registry of Digital Masters,<sup>11</sup> ali katerega drugega projekta digitalnega arhiviranja. Uporaba polj 506, 533, 538 in 583 je dovoljena le, kadar gre za zapise arhivskih digitalnih originalov.

Največje razlike v primerjavi s prejšnjim načinom obdelave se pojavljajo pri uporabi polj 020, 260, 300, 533, 710, 730, 776 in 856. V originalnem (od dobavitelja neodvisnem) bibliografskem zapisu se ne uporabljajo opombe in točke dostopa, ki izhajajo iz določene verzije gradiva (opombe o sistemskih zahtevah, reprodukciji, dobavitelju in imenu paketa ter opombe, ki se nanašajo na URL-dostop, vezan na določeno knjižnico).

Kljub novim priporočilom pa težave z upravljanjem lokalnih paketov zapisov ostajajo. Zelo malo dobaviteljev je namreč obstoječe zapise konvertiralo v "nevtralno" obliko. Še vedno se razširja na tisoče zapisov različnih dobaviteljev, ki vsebujejo opombe o reprodukciji, ime paketa in dobavitelja, ali pa so napačno katalogizirani kot izvirne digitalne izdaje (angl. *born-digital editions*).

Velika ovira pri uporabi priporočil na lokalnem nivoju je pomanjkanje zanesljivega identifikatorja za avtomatizirano kolokacijo enakovrednih pojavnih oblik. Priporočila nekako predpostavljajo, da bo katalogizator med seboj primerjal različne datoteke ali datoteko z metapodatki drugega katalogizatorja. Še vedno pa ni jasno, kako naj knjižnice, ki dobivajo zapise za e-knjige iz veliko različnih virov, učinkovito identificirajo in združijo podvojene zapise.

Dobavitelji ponujajo zelo različne zapise:

- nekateri se trdno držijo katalogizacijskih standardov;
- drugi poskrbijo za osnovne zapise MARC šele po dolgih pogajanjih;
- nekateri pa poskrbijo za nenavadne dodatke, ki so omejeni le na manjše število zapisov.

Katalogizatorji v UHL pred uvozom zapisov ocenijo vsak paket in se nato odločijo, kako s čim manj dela in kasnejšega vzdrževanja dopolniti lokalni katalog.

Posamezna obdelava e-knjig je zelo redka, največkrat samo v primerih, ko:

- je delo zelo pomembno za zbirko knjižnice in je vidnost v WorldCat prva prioriteta,
- je delo stalno dostopno zaradi enkratnega nakupa,
- dobavitelj ne zagotavlja zapisov ali pa so ti zelo slabe kakovosti,
- je enota samostojno delo v odprtem dostopu.

Pri tem je potrebna posebna previdnost pri delih, ki so v tiskani obliki končna, v spletnem okolju pa delujejo kot integrirni viri v elektronski obliki, kot npr. enciklopedije, ki se v tiskani obliki pojavljajo kot monografije, v digitalni obliki pa gre za spletne strani in portale.

Za velike pakete zapisov, najpogosteje za določene zbirke, na katere so naročene knjižnice, ki redno dodajajo in opuščajo naslove, se je izkazalo, da je najbolj učinkovito, če vsakič star paket zapisov enostavno "prekrijejo" z novim. Pri tem so koristi hitrega dostopa in enostavnost dopolnjevanja močno prevladale nad morebitnimi pomanjkljivostmi in napakami v zapisih, ki jih zagotavljajo dobavitelji.

V primeru zapisov za druge (nemonografske) oblike del, kot so poezija, kratke zgodbe, različna poročila in primarni viri (pisma, intervjuji), so za razlikovanje od "tradicionalnih" knjig v besedilu povezave začeli uporabljati besedo "e-besedilo". Pri tem pa katalogizatorji ne iščejo posameznih naslovov, pri katerih bi uporabili to oznako, temveč jo uporabijo pri celotnem paketu zapisov, kadar od dobavitelja dobijo podatke, da zbirko sestavljajo

pretežno literarna dela, primarni viri in druge neknjižne vsebine. Ob vse večji uporabi fasetnega brskanja v katalogih so začeli razmišljati o uvedbi podrobnejše določitve formata za posamezne vrste gradiva.

Pred sprejemom novih priporočil PCC je bila glavna bibliografska značilnost e-knjige, da je imela status reprodukcije. Skladno z LC Rule Interpretations 1.11A je bila reproducirana e-knjiga obravnavana kot sekundarna pojavna oblika neelektronskega izvornika. Območje založništva in fizični opis sta se nanašala na izvornik, polje 533 (elektr. reprodukcija) pa je služilo za opis reprodukcije. Izvirno digitalna e-knjiga se je obravnavala kot enkratna pojavna oblika. Območje založništva je vsebovalo informacije o elektronski verziji, večina takšnih zapisov pa ni vsebovala nobenega fizičnega opisa.

Nova priporočila v veliki meri odpravljajo te razlike. Katalogizator sicer lahko naredi nov zapis v primeru bistvenih razlik med elektronsko in tiskano obliko, vendar priporočila pojmujejo enakovredne pojavne oblike zelo široko in odsvetujejo kreiranje ločenih zapisov.

Priporočila več ne zahtevajo ločevanja zapisov za elektronske in tiskane izdaje, ali standardizacije opomb o reprodukciji, ter točk dostopa za različne pakete in dobavitelje. Namesto tega je treba zapise za reprodukcije, elektronske pojavne oblike, izdane skupaj s tiskanimi deli, ter izvorna digitalna dela podrobno preveriti in zagotoviti, da ne vsebujejo opuščenih polj.

Za lažje upravljanje z zapisi katalogizatorji UHL priporočajo:

- zbiranje in hranjenje vzorčnih zapisov,
- seznam polj, ki morajo (ali ne smejo) biti v zapisu,
- zagotavljanje trajnih podatkov za polja, ki zahtevajo enolično kodiranje (da bi zmanjšali število napak pri vnosu podatkov).

Proces paketnega urejanja zapisov poteka s pomočjo programa MarcEdit, ki omogoča hkratno urejanje več tisoč zapisov hkrati, v "čitljivi" besedilni obliki (txt). Pri upravljanju z velikim številom zapisov ima polje 001, ki vsebuje enolični identifikator zapisa, poseben pomen v kasnejšem procesu upravljanja in vzdrževanja zapisov za določeni paket. Zato je že na začetku paketnega urejanja treba nujno preveriti, ali ima vsak zapis svoj identifikator. Za ta namen se v MarcEdit uporablja funkcija "Field Count" – po štetju polj se mora število zapisov v paketu ujemati s številom polj 001. Če se številki ujemata, se v naslednjem koraku uporabi funkcija "Record Deduplication", ki preveri, ali v paketu zapisov obstajajo kakšni podvojeni zapisi. Ker je paketno procesiranje in nalaganje zapisov močno odvisno od pravilne sintakse

in vsebine polj, se po zaključenem urejanju s funkcijo "Validate" programsko preveri, ali so morda ostale še kje kakšne napake. Nato se zapisi pretvorijo nazaj v obliko MARC (mrc) in naložijo v katalog (lokalno bazo podatkov).

Pomemben del upravljanja z zapisi za e-knjige je vzdrževanje in dopolnjevanje paketov. Vsak paket ima svoj "urnik" dopolnjevanja, ki se razlikuje od paketa do paketa. Nekateri dobavitelji pripravljajo mesečne datoteke dopolnitev in sprememb, nekateri ne, in kadar je knjižnica naročena na več različnih paketov, lahko postane njihovo obvladovanje zelo zahtevno. Če dobavitelj ne zagotavlja mesečnih datotek, je treba vsakič znova naložiti celoten paket in nato poiskati zapise, ki ob tem niso bili "prekriti", in jih pobrisati.

Ker večina dobaviteljev ne skrbi za ažurno obveščanje o dostopnosti novih zapisov ali paketov, je potreben reden nadzor nad delom katalogizatorjev. Zaradi zapletenosti procesa v primeru množice paketov je nujno natančno in ažurno vzdrževanje spremne dokumentacije. Katalogizatorji UHL priporočajo: ime paketa in dobavitelja (s spletnim naslovom in geslom za dostop do paketov z zapisi), sintakso enoličnega identifikatorja zapisa, ki jo uporablja dobavitelj, datum zadnjih dopolnitev, podatek o pogostnosti dopolnjevanja, ter seznam polj, ki so v posameznem paketu bistvena za njegovo učinkovito vzdrževanje.

Paketno upravljanje bibliografskih zapisov je dejavnost, ki presega tradicionalno katalogizacijo in sega na področje dobaviteljevih postopkov za objavo metapodatkov, modelov nabave in sistemskih vidikov upravljanja s podatki. Po podrobni in z več vidikov poučni predstavitvi bibliografskih postopkov avtorici ugotavljata, da je po določitvi standardov za kreiranje zapisov nujno potreben korak v smeri določitve in poenotenja postopkov v vsej dobavni verigi bibliografskih podatkov za e-knjige. Šele to bo knjižnicam omogočilo, da uporabnikom v svojih katalogih zagotovijo enostaven, konsistenten in ažuren dostop do e-knjig.

Knjižničarska skupnost naj si skupaj z uporabniki, avtorji in založniki prizadeva za ureditev vseh področij dostopa in uporabe e-knjig. Za katalogizatorje to pomeni uskladitev zahtev za vsebino bibliografskih zapisov za e-knjige in standardizacijo postopkov za njihovo upravljanje. Jasne zahteve bi okrepile pogajalska izhodišča v procesu nabave in od založnikov bi lahko zahtevali več za svoje uporabnike. Ta več vključuje tudi standardiziran bibliografski zapis ustrezne kakovosti za e-knjige in njihove zbirke, ki v spletnem okolju predstavlja neke vrste rojstni list, kar bi pospešilo in

olajšalo dostop ter zmanjšalo stroške upravljanja. Morda bi se na račun tega lahko povečale zbirke e-knjig, ki jih imajo knjižnice na voljo za svoje uporabnike. Žal pa se založniki kljub odločno previsokim cenam storitev ne zmenijo kaj prida za tegobe uporabnikov in knjižničarjev, morda malo bolj poskrbijo le za avtorje. Knjižničarska skupnost bi morala v primeru e-knjig trdneje zastopati interese uporabnikov in avtorjev, sicer bodo založniki še naprej povzročali nepotrebne dodatne stroške in ovire pri dostopu.

## Opombe

- 1 <http://bit.ly/a8dDKF> (vsi spletni naslovi preverjeni 12. 9. 2010)
- 2 <http://bit.ly/aaDPby>
- 3 <http://bit.ly/9vplm4>
- 4 <http://bit.ly/c0FneQ>
- 5 <http://bit.ly/bFI2gL>
- 6 <http://info.lib.uh.edu/>
- 7 <http://www.loc.gov/catdir/pcc/>
- 8 <http://www.loc.gov/catdir/pcc/bibco/PN-Guide.pdf>
- 9 <http://people.oregonstate.edu/~reaset/marcedit/html/index.php>
- 10 <http://www.loc.gov/marc/bibliographic/ecbdhome.html>
- 11 <http://www.oclc.org/digitalregistry/>

Igor Brbre