

Διαχείριση τεκμηρίων στο μεταβαλλόμενο περιβάλλον πληροφόρησης

Μιχάλης Γερόλιμος

Τμήμα Αρχειονομίας – Βιβλιοθηκονομίας

mike@ionio.gr

Ανέστης Σίτας

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

sitas@lit.auth.gr

Ράνια Κώνστα

Τμήμα Αρχειονομίας-Βιβλιοθηκονομίας, Ιόνιο Πανεπιστήμιο

rania@ionio.gr

Περίληψη

Αναφέρονται πληροφορίες που αφορούν την περιγραφή και τη διαχείριση υλικού βιβλιοθηκών. Παρουσιάζονται, συνοπτικά, οι προσπάθειες δημιουργίας νέων και βελτίωσης των παραδοσιακών/συμβατικών προτύπων καταχώρησης και περιγραφής, όσον αφορά όλα τα είδη των τεκμηρίων (χαρτώα, ηλεκτρονικά και ψηφιακά).

Βιβλιογραφική περιγραφή, κανόνες καταλογογράφησης και ISBDs

Η βιβλιογραφική περιγραφή ενός δημοσιεύματος έχει ως στόχο την καταγραφή όλων των στοιχείων που αποδίδουν μοναδική ταυτότητα σε κάθε τεκμήριο και καθορίζουν τα σημεία πρόσβασης, τα οποία κάνουν δυνατό τον εντοπισμό του. Η δημιουργία της βιβλιογραφικής εγγραφής δημιουργείται στη βάση καθιερωμένων κανόνων, διατάξεων και οδηγιών. Το κύριο εργαλείο βιβλιογραφικής περιγραφής είναι οι κανόνες καταλογογράφησης. Αν και οι πρώτες προσπάθειες δημιουργίας τους ξεκίνησαν στα μέσα του 19^{ου} αιώνα, μόλις το 1967 παρουσιάστηκαν οι **AACR** (Anglo-American Cataloging Rules) ως πρότυποι κανόνες που αφορούσαν την καταλογογράφηση των τεκμηρίων των βιβλιοθηκών. Η τρέχουσα έκδοσή τους είναι οι AACR2, *2002 Revision* (με προσθήκες το 2003, 2004 και το 2005). Έχουν συμπεριλάβει κανόνες που αφορούν την περιγραφή των πηγών του Διαδικτύου και άλλα είδη πόρων χωρίς, ωστόσο, να αποτελούν έναν εξ ολοκλήρου νέο κώδικα. Το 2009 αναμένεται να δημοσιευτεί η νεότερη έκδοσή τους με τίτλο: «RDA - Resource Description and Access» (AACR JCS, 2007).

Η ανάγκη για τυποποίηση των διεργασιών της καταλογογράφησης σε διεθνή κλίμακα, οδήγησε την IFLA, από το 1971, στη δημιουργία μιας σειράς προτύπων βιβλιογραφικής περιγραφής, γνωστά ως International Standard Bibliographic Descriptions (ISBDs). Ορίζουν τη δομή που πρέπει να έχει μια βιβλιογραφική εγγραφή, τη σειρά των στοιχείων της περιγραφής και το σύστημα στίξης, γεγονός που δίνει τη δυνατότητα ανταλλαγής των βιβλιογραφικών εγγραφών και τη μετατροπή τους σε μορφή που μπορεί να γίνει κατανοητή από υπολογιστή (AACR2 Homepage, 2006). Η ενσωμάτωση των ISBDs στη δομή των AACR2 αποτέλεσε μια από τις σημαντικότερες εξελίξεις του χώρου της διαχείρισης τεκμηρίων.

Ενώ οι συλλογές των βιβλιοθηκών τους τελευταίους αιώνες απαρτίζονταν, ουσιαστικά, από χαρτώο υλικό (βιβλία, περιοδικά κ.λπ.) με την πάροδο του χρόνου εμφανίζονται νέα σε μορφή τεκμήρια που οδηγούν στη δημιουργία νέων εργαλείων, ώστε να είναι δυνατή η βιβλιοθηκονομική επεξεργασία τους. Η IFLA, οδηγούμενη από την αυξανόμενη ανάγκη για ένα ξεχωριστό ISBD που αφορά ειδικά τα αρχεία υπολογιστών, δημοσίευσε το 1990 το ISBD (CF-Computer Files). Οι ηλεκτρονικοί πόροι όμως είναι προϊόντα μιας τεχνολογίας που εξελίσσεται με πολύ γρήγορο ρυθμό. Ως αποτέλεσμα αυτής της εκτίμησης, αποφασίστηκε η επανεξέταση και η αναθεώρηση του ISBD (CF), που οδήγησε το 1997 στην έκδοση του ISBD (ER - Electronic Resources) (IFLANET, Family of ISBDs, 2005).

Εργαλεία θεματικής πρόσβασης

Η απόδοση θεματικών όρων στα τεκμήρια αποτελεί προσπάθεια «κωδικοποίησης» του περιεχομένου τους, ώστε να είναι πιο εύκολη η αναζήτησή τους από το χρήστη. Για την απόδοσή τους έχουν δημιουργηθεί διάφορα εργαλεία, σκοπός των οποίων είναι η σύνταξη ελεγχόμενων λεξιλογίων απόδοσης θεματικών όρων. Τέτοια εργαλεία είναι, για παράδειγμα, οι θησαυροί και οι θεματικές επικεφαλίδες, τα πιο διαδεδομένα εκ των οποίων είναι οι Library of Congress Subject Headings (LCSH), οι MeSH (Medical Subject Headings) και το Eurovoc.

Αν και η δημιουργία και η χρήση των **LCSH** ξεκίνησε το 1897, η βασική έντυπη 9^η έκδοσή τους παρουσιάστηκε το 1980, ενώ τώρα βρίσκονται στην 30^η έκδοσή τους. Από το 1995 είναι διαθέσιμες μέσω του Internet και διατίθενται μέσω του *Classification Web*.¹ Η πρώτη έντυπη έκδοση των **MeSH** κυκλοφόρησε το 1954, ενώ ήδη η πρόσβαση παρέχεται δωρεάν από το MeSH Website.² Τέλος το **EUROVOC** είναι ένας πολύγλωσσος θησαυρός των οργάνων της Ευρωπαϊκής Ένωσης που πρωτοδημοσιεύτηκε το 1984. Η τρέχουσα έκδοση διατίθεται στις 16 επίσημες γλώσσες της Ευρωπαϊκής Ένωσης (μεταξύ των οποίων και η ελληνική) και η πρόσβαση είναι ελεύθερη στην ιστοσελίδα του *Eurovoc thesaurus*.³

Στο έργο της θεματικής περιγραφής των τεκμηρίων συνεισφέρουν και τα διάφορα ταξινομικά συστήματα, όπως το DDC, το LCC, το UDC. Από τα σύγχρονα ταξινομικά συστήματα ευρέως διαδεδομένο είναι αυτό Dewey Decimal Classification που η ανάπτυξή του ξεκίνησε το 1876 και το Library of Congress Classification το οποίο κυρίως αναπτύχθηκε μεταξύ του 1899-1940.

Μέσα δημοσίευσης εργαλείων

Το συμβατικό μέσο δημοσίευσης όλων των εργαλείων είναι η έντυπη μορφή. Με την εξέλιξη της τεχνολογίας τα εργαλεία αυτά άρχισαν να είναι διαθέσιμα και σε άλλες μορφές. Αρχικά μεταφέρθηκαν σε μικροφόρμες, αλλά κυρίως σε CD-ROM, τα οποία έδιναν και τη δυνατότητα εγκατάστασής τους σε τοπικά δίκτυα. Η πιο γνωστή σχετική υπηρεσία είναι το **Cataloger's Desktop**, το οποίο εκτός των κανόνων καταλογογράφησης, παρέχει πρόσβαση στα σημαντικότερα εργαλεία υλοποίησής της. Όμως η μορφή που επικρατεί σταδιακά είναι η πρόσβαση μέσω Web. Η Library of Congress έπαυσε την έκδοσή σε CD-ROM με το issue 4 (2005) και η πρόσβαση γίνεται πλέον μόνο από την ειδική Web version⁴, αλλά και μέσω του **Classification Web**, απ' όπου δίνεται πρόσβαση σχεδόν στο σύνολο των υπηρεσιών και των εργαλείων που απαιτούνται (ταξινομικοί πίνακες, θεματικοί όροι, authority files κ.λπ.). Η μορφή αυτή δίνει περισσότερες και πιο εξελιγμένες δυνατότητες.

Καταλογογράφηση αναγνώσιμη από μηχανή: MARC

Τα προγράμματα μηχανοργάνωσης των βιβλιοθηκών έχουν την ανάγκη εφαρμογής ενός προτύπου για την αυτοματοποιημένη ανάγνωση των δεδομένων. Τη δυνατότητα αυτή δίνουν τα MARC formats. Η δημιουργία τους ξεκίνησε από τη Library of Congress των ΗΠΑ κατά τη διάρκεια της δεκαετίας του 1960 με τη δημιουργία του LC-MARC (USMARC) format. Η μορφή αυτή μαζί με το ομοειδές CAN-MARC εναρμόνισαν τις διαφορές τους, με αποτέλεσμα στις αρχές του 1999 να εκδοθεί ένας τόμος με την ονομασία MARC21. Εν τω μεταξύ είχαν αναπτυχθεί από άλλες βιβλιογραφικές υπηρεσίες πάνω από 20 διαφορετικά MARC πρότυπα όπως το UKMARC. Το πρόβλημα των πολλών και διαφορετικών προτύπων, της ανταλλαγής τους και της συμβατότητάς τους προσπάθησε να λύσει το 1977 η IFLA με την έκδοση του UNIMARC.

Από το 1981 υιοθετήθηκε το **ISO 2709**, το οποίο αντικαταστάθηκε από το ISO 2709:1996 «Information and documentation - Format for Information Exchange». Πρόκειται για ένα

¹ LC, Classification Web, <http://classificationweb.net>

² NLM, MeSH Browser, <http://www.nlm.nih.gov/mesh/MBrowser.html>

³ Eurovoc thesaurus, <http://europa.eu.int/celex/eurovoc>

⁴ Cataloger's Desktop, <http://www.loc.gov/cds/desktop>

διεθνές σχήμα ανταλλαγής δεδομένων, με στόχο να παράσχει ένα διεθνώς αποδεκτό πρότυπο στο οποίο μπορούν να προσαρμοστούν όλα τα υπάρχοντα formats.

Μια από τις σημαντικότερες και πιο ενδεικτικές ενέργειες προσαρμογής του MARC στα νέα μέσα δημοσίευσης των τεκμηρίων, ξεκίνησε από το 1993 και αφορά τη δημιουργία ειδικού πεδίου που περιέχει πληροφορίες, οι οποίες είναι αναγκαίες για τον εντοπισμό και την πρόσβαση σε ηλεκτρονικούς πόρους, με τη χρήση του πεδίου 856 «Electronic Location and Access» στα USMARC formats. Αμέσως μετά το πεδίο αυτό υιοθετήθηκε και από το UNIMARC.

Περιγραφή ψηφιακών αντικειμένων: μεταδεδομένα

Η συνεχής αύξηση των ψηφιακών αντικειμένων (αρχεία, εικόνες, ήχος, προγράμματα υπολογιστών κ.λπ.) που είναι διαθέσιμα μέσω του Web, οδήγησε τις βιβλιοθήκες στην ενσωμάτωσή τους στις συλλογές τους. Έχει ήδη αρχίσει να συζητείται η μετατόπιση του ενδιαφέροντος από την αυστηρή τήρηση δύσκαμπτων βιβλιοθηκονομικών κανόνων, σε πρότυπα που μπορούν να υποστηριχθούν από αυτοματοποιημένα συστήματα οργάνωσης συμβατικών ή «ψηφιακών» βιβλιοθηκών.

Για τη διαχείριση τεκμηρίων παραδοσιακής μορφής κρίνεται επαρκής, ως ένα βαθμό, η χρήση εργαλείων που για πολλές δεκαετίες χρησιμοποιήθηκαν για την επεξεργασία του υλικού των βιβλιοθηκών. Όμως τα δεδομένα όπως ο συγγραφέας, ο τίτλος και οι θεματικοί όροι, αποδεικνύονται συχνά ανεπαρκή για πολλούς τύπους υλικών, κυρίως δε όσον αφορά τα ψηφιακά αντικείμενα (Hunt and Ethington, 1997). Στις συμβατικές βιβλιοθήκες η δημιουργία των βιβλιογραφικών εγγραφών ήταν έως ένα σημείο ελεγχόμενη. Στο Διαδίκτυο παρατηρείται έλλειψη οργάνωσης ελέγχου των διαθέσιμων εργαλείων έρευνας και των μηχανών αναζήτησης. Το γεγονός αυτό οδηγεί στην ανάγκη λειτουργίας βάσεων δεδομένων που περιέχουν σύνθετες βιβλιογραφικές περιγραφές που βασίζονται σε διαφορετικά πρότυπα, ανάλογα με τον τύπο και το είδος του αντικειμένου που περιγράφουν. Μια από τις πιο ενδιαφέρουσες προσεγγίσεις είναι η χρήση μεταδεδομένων για την περιγραφή των πόρων στο Διαδίκτυο. Ως μεταδεδομένα ορίζονται τα δεδομένα που περιγράφουν τις ιδιότητες μιας πηγής, χαρακτηρίζουν τις σχέσεις της, υποστηρίζουν την ανακάλυψη και την αποτελεσματική της χρήση και υπάρχουν σε ένα ηλεκτρονικό περιβάλλον (Velluci, 1999).

Dublin Core Metadata Element Set

Η ανάγκη ύπαρξης ενός κοινού προτύπου περιγραφής και απόδοσης μεταδεδομένων από όλες τις ενδιαφερόμενες κοινότητες οδήγησε στη δημιουργία προτύπων που συνδυάζουν τις ανάγκες όσο το δυνατόν περισσότερων κοινοτήτων, αλλά και ενός προτύπου που να είναι ταυτόχρονα απλό και εφαρμόσιμο, χωρίς να απαιτεί υψηλή ειδικευση. Έτσι οδηγηθήκαμε στη δημιουργία του Dublin Core που ξεκίνησε το 1995 από το OCLC σε συνεργασία με άλλους οργανισμούς. Στόχος του είναι η παράλληλη, με άλλα σχήματα, λειτουργία και η δυνατότητα εκμετάλλευσης από κάθε δυνατή χρήση της τεχνολογίας περιγραφής και κωδικοποίησης πληροφοριών. Μπορεί να κωδικοποιηθεί στις γλώσσες HTML ή XML, που γνωρίζουν πλατύτερη εφαρμογή από τις μορφές MARC, οι οποίες είναι περιορισμένες στην κοινότητα των βιβλιοθηκών.

Νέα εργαλεία οργάνωσης και σύνταξης μεταδεδομένων

Τα MARC formats αποτελούν τα θεμέλια της αυτοματοποίησης των βιβλιοθηκών. Η εμφάνιση, όμως, το 1996 της γλώσσας XML και του RDF (Resource Description Framework) αποτελεί μια δυναμική εξέλιξη που δίνει νέες δυνατότητες προγραμματισμού. Η XML και το RDF αυτά καθ' αυτά, δεν είναι μορφές μεταδεδομένων, αλλά μπορούν να χρησιμοποιηθούν για οποιοδήποτε εφαρμογή. Οι προσπάθειες μετατροπής των εγγραφών MARC σε XML έχουν αρχίσει από τα τέλη της δεκαετίας του 1990. Το Library of Congress Network Development και το MARC Standards Office αναπτύσσουν το MARCXML. Η LC από το 1995 δημιούργησε DTDs που επέτρεπαν τη μετατροπή δεδομένων MARC σε SGML. Αργότερα αυτό το SGML DTD μετατράπηκε σε XML DTD, το οποίο είναι διαθέσιμο για δοκιμαστική χρήση από το Μάρτιο του 2003. Συμπληρωματικά, το 2004 η IFLA σε συνεργασία με την National Library of Portugal και

την BookMARC δημιούργησαν ένα πρωτότυπο του UNIMARC manual σε XML (Carvalho, 2005).

MODS & METS

Το MODS⁵ (Metadata Object Description Schema) είναι συμβατό με την XML και το XMLMARC με στόχο να φέρει επιλεγμένα δεδομένα από τις υπάρχουσες εγγραφές MARC21 και να επιτρέψει τη δημιουργία πρωτογενών εγγραφών περιγραφής πηγών. Είναι ιδιαίτερα χρήσιμο για τους ηλεκτρονικούς πόρους στις περιπτώσεις όπου απαιτείται οι εγγραφές να μεταφερθούν μέσω των πρωτοκόλλων XML. Το 2002 έγινε δημόσια διαθέσιμο για δοκιμαστική χρήση. Το METS⁶ (Metadata Encoding & Transmission Standard) παρέχει ένα XML document format για την κωδικοποίηση των μεταδεδομένων που απαιτούνται για τη διαχείριση ψηφιακών αντικειμένων των βιβλιοθηκών εντός ενός αποθετηρίου (repository) και για την ανταλλαγή τους μεταξύ διαφορετικών αποθετηρίων (ή μεταξύ των αποθετηρίων και των χρηστών τους). Δεν περιέχει περιγραφικά μεταδεδομένα, για την απόδοση των οποίων μπορεί να χρησιμοποιηθούν ποικίλα πρότυπα, συνήθως DC και MODS.

Κωδικοποίηση πλήρους κειμένου

Η βιβλιογραφική περιγραφή όσο και αν είναι αντιπροσωπευτική του τεκμηρίου, συνεχίζει να μην παρέχει το ουσιαστικό ζητούμενο για ένα χρήστη: την ουσιαστική πρόσβαση στην πληροφορία. Για το λόγο αυτό η επιστημονική κοινότητα οδηγήθηκε στην αναζήτηση κωδικοποίησης του πλήρους κειμένου. Το **TEI (Text Encoding Initiative)**⁷ είναι μια προσπάθεια, που ξεκίνησε το 1987 και έχει εξελιχθεί σε ένα ισχυρό μέσο αναζήτησης, ευρετηρίασης και αποθήκευσης πληροφοριών. Είναι ήδη συμβατό με τη γλώσσα XML. Το TEI header και η εγγραφή MARC μπορεί να υπάρξουν παράλληλα, καθώς δεν αντικαθιστούν το ένα το άλλο και επιτρέπεται η δημιουργία μιας σύνδεσης μεταξύ του τεκμηρίου TEI και της αντίστοιχης εγγραφής MARC (McCallum, 2004).

Αυτόματη εξαγωγή μεταδεδομένων

Το μεγαλύτερο μέρος της βιβλιοθηκονομικής κοινότητας υποστηρίζει με μεγάλη θέρμη την άποψη πως οι βιβλιοθήκες είναι σε θέση να δημιουργούν τα μεταδεδομένα του συνόλου του υλικού τους και να τα καθιστούν διαθέσιμα στο Web χωρίς να θυσιάζονται τα απαραίτητα περιγραφικά πρότυπα. Το κυριότερο πρόβλημα στον εντοπισμό των πηγών του Web παραμένει η έλλειψη της χρήσης αποτελεσματικών ελεγχόμενων λεξιλογίων (όπως DDC, LCC, LCSH) θεματικής πρόσβασης. Η αυτόματη παραγωγή μεταδεδομένων αποτελεί αντικείμενο εκτεταμένης έρευνας κυρίως λόγω της μείωσης του κόστους που επιφέρει η εφαρμογή της. Παρόλα αυτά δεν πρέπει να διαφεύγει της προσοχής μας πως (τουλάχιστον προς το παρόν) αρτιότερα αποτελέσματα επιτυγχάνονται μόνο με τη συμπληρωματική εφαρμογή μη αυτοματοποιημένων και αυτοματοποιημένων μεθόδων (The AMeGA project, 2006).

Συμπεράσματα

Για την αποτελεσματική ενσωμάτωση και διαχείριση των πηγών του Διαδικτύου από τις βιβλιοθήκες είναι αναγκαία η εφαρμογή απλών οργανωτικών δομών και η παρουσίασή τους με έναν τρόπο ώστε να είναι χρήσιμες και κατανοητές στους χρήστες των βιβλιοθηκών, συμβατικών ή ψηφιακών. Είναι λογικό να υποθέσουμε ότι κάποια στιγμή θα υπάρξει ένα εξ ολοκλήρου, τυπικά αλλά και ουσιαστικά, ψηφιακό περιβάλλον διαχείρισης της πληροφορίας. Όμως για ένα αδιευκρίνιστο χρονικό διάστημα, το πραγματικό πληροφοριακό περιβάλλον θα

⁵ The Library of Congress, *MODS, Metadata Object Description Schema official website*, <http://www.loc.gov/standards/mods/>

⁶ The Library of Congress, *METS, Metadata Encoding & Transmission Standard official website*, <http://www.loc.gov/standards/mets/>

⁷ The Text Encoding Initiative, <http://www.tei-c.org/>

είναι υβριδικό, δηλαδή αυτό μέσα στο οποίο συνυπάρχουν και συλλειτουργούν θεσμοί, εργαλεία και συνήθειες από δύο κόσμους, τον παραδοσιακό και τον ψηφιακό. Πρέπει, λοιπόν, να είμαστε σε θέση να εκμεταλλευτούμε τα πλεονεκτήματα αλλά και τις δυνατότητες που αυτό προσφέρει. Κρατώντας ό,τι θεωρούμε πως πρέπει να μείνει σταθερό, προσαρμόζοντας όσα έχουν τη δυνατότητα αυτή, καταργώντας παλιές συνήθειες και εργαλεία, όταν τα παλιά δεν είναι σε θέση να ικανοποιήσουν τις νέες ανάγκες, αλλά ταυτόχρονα δημιουργώντας νέα εργαλεία όταν η ανάγκη προβάλλει επιτακτικά.

Βιβλιογραφία

1. AACR JCS (2007) *A brief history of AACR*, διαθέσιμο στο <http://www.collectionscanada.ca/jsc/history.html>
2. AACR2 Homepage (2006), <http://www.aacr2.org/access.html>
3. The AMeGA project, *Final report* (2006), http://www.loc.gov/catdir/bibcontrol/lc_amega_final_report.pdf
4. Carvalho, Joaquim (2005). An XML representation of the UNIMARC Manual: a working prototype, *71st IFLA General Conference*, διαθέσιμο στο <http://www.ifla.org/IV/ifla71/papers/199e-Carvalho.pdf>
5. Hunt, L. (Li) and Ethington, P.J. (Nov. 1997), "The utility of spatial and temporal organization in digital library construction", *The Journal of Academic Librarianship*, Vol. 23, pp. 475-483.
6. IFLANET, *Family of ISBDs* (2005), διαθέσιμο στο <http://www.ifla.org/VI/3/nd1/isbdlist.htm>
7. McCallum, S. (2004), *Recent Developments in MARC with XML*, διαθέσιμο στο <http://www.loc.gov/acq/conser/MARC-overview.html>
8. The Library of Congress, *MARC standards*, <http://www.loc.gov/marc>
9. The Library of Congress, *MARC XML, MARC21 XML Schema official website*, <http://www.loc.gov/standards/marcxml>
10. Velluci, S. (1999), "Metadata", *Annual Review of Information Science and Technology*, Vol. 33, pp. 187-222.