

**Οι ακαδημαϊκές βιβλιοθήκες ως πάροχοι δεδομένων στο σημασιολογικό ιστό :
η μετάβαση από τη μία εγγραφή ανά πόρο στα δίκτυα σχέσεων μεταξύ
αυτόνομων οντοτήτων**

Μανόλης Πεπονάκης

Εθνικό Ίδρυμα Ερευνών / Εθνικό Κέντρο Τεκμηρίωσης

epero@ekt.gr

Περίληψη

Η μελέτη εστιάζει στην κρίσιμη καμπή στην οποία βρίσκονται οι βιβλιοθήκες οι οποίες από διαχειριστές των τεκμηρίων των συλλογών τους καλούνται να γίνουν διαχειριστές της γνώσης που συνδέεται με κάθε είδους τεκμήρια εντός και εκτός των τειχών τους. Η εργασία χωρίζεται σε δύο βασικά μέρη.

Στο πρώτο μέρος καταδεικνύονται οι αλλαγές που συντελούνται στο χώρο οργάνωσης της πληροφορίας. Από τη μία πλευρά η ανάπτυξη του σημασιολογικού ιστού (semantic web) και των συνδεδεμένων δεδομένων (linked data), έχοντας ως γλώσσα έκφρασης το Resource Description Framework (RDF), αλλάζει τις βασικές υποδομές αλλά και τις λογικές μέσω των οποίων μπορούν να εκφραστούν πλέον τα μεταδεδομένα. Από την άλλη πλευρά οι Functional Requirements of Bibliographic Records (FRBR) καταργούν τη λογική της επίπεδης περιγραφής όπου υπήρχε μία εγγραφή ανά πόρο και πλέον η περιγραφή του πόρου είναι ένα σύνολο οντοτήτων οι οποίες συνδέονται μεταξύ τους. Συνδυάζοντας την εκφραστικότητα που παρέχει το RDF με τη σημασιολογία των FRBR οι βιβλιοθήκες βρίσκονται σε μια εποχή όπου καλούνται να αναθεωρήσουν πολλές από τις πρακτικές τους.

Στο δεύτερο μέρος τεκμηριώνεται πώς η εφαρμογή των παραπάνω μέσω της -ήδη δρομολογημένης- μετάβασης προς το σημασιολογικό ιστό έχει σημαντικές επιπτώσεις στην καταγραφή και κωδικοποίηση των δεδομένων. Στην ουσία επιφέρει αλλαγή του ρόλου των ακαδημαϊκών βιβλιοθηκών μετατρέποντάς τις σε παραγωγούς και εκδότες σημασιολογικής πληροφορίας. Σε αυτή την ενότητα εξετάζεται ο ρόλος των ακαδημαϊκών βιβλιοθηκών ως φορέων που μπορούν να εξασφαλίσουν την ακεραιότητα (integrity) της πληροφορίας και μέσω της πιστοποιημένης προέλευσής (provenance) της να αυξηθεί η εμπιστοσύνη (trust) των τρίτων σε αυτήν. Η εξασφάλιση των τριών αυτών στοιχείων (ακεραιότητα, προέλευση, εμπιστοσύνη) καθιστά την παραγόμενη πληροφορία αξιόπιστη βάση ώστε τρίτοι να επενδύσουν σε αυτήν είτε επαναχρησιμοποιώντας την αυτούσια είτε "χτίζοντας" πάνω σε αυτήν.

Λέξεις κλειδιά

Σημασιολογικός ιστός, Ακαδημαϊκές βιβλιοθήκες, Δημοσίευση σημασιολογικής πληροφορίας, Φορμαλιστική αναπαράσταση, Βιβλιοθήκες ως πάροχοι πληροφορίας

Summary

The study examines the critical juncture where libraries from managers of their collections have to become managers of the knowledge associated with all sorts of resources within and outside their walls. The work is divided into two major parts.

In the first part of the study we deploy the changes taking place in the field of information organization. On one hand we find the development of the semantic web and linked data, having as expression language the Resource Description Framework (RDF), which changes the basic infrastructure and the ways through which metadata could be expressed. On the other hand the Functional Requirements of Bibliographic Records (FRBR) discard the theory and practice of flat records where there is one record per resource. Thus the description of the resource becomes a set of entities which are interconnected. Libraries are before the need to revise many of their practices by combining the expressiveness of RDF with the semantics of FRBR.

The second part of the study documents how the –already routed– transition to the semantic web and the corresponding implementation of the above have a major impact to the recording and coding of data. Actually, it changes the role of academic libraries by converting them to producers and providers of semantic information. This section focuses on the role of academic libraries as institutions that can ensure the integrity of information through its certified provenance which provides trust to third parties in order to either re-use this exact information or to build upon it.

Keywords

Semantic web, Academic libraries, Semantic information publishing, Formal representation, Libraries as information providers

Εισαγωγή

Σε όλη τη διάρκεια του 20^{ου} αιώνα οι κατάλογοι των βιβλιοθηκών αντιμετώπιζαν το φυσικό αντικείμενο του βιβλίου ως το αδιαίρετο στοιχείο που αποτελούσε το σημείο αναφοράς για κάθε εγγραφή του καταλόγου. Ο αδιαίρετος χαρακτήρας του φυσικού αντικειμένου, βιβλίου κατά κανόνα (Smiraglia, 2009), οδηγούσε με τη σειρά του στη λογική πως μία βιβλιογραφική εγγραφή περιγράφει ένα τεκμήριο και ένα τεκμήριο περιγράφεται από μία εγγραφή. Δηλαδή υπήρχε ανάμεσα σε πόρους και εγγραφές μια αναλογία ένα προς ένα. Είναι γεγονός πως όσο η βιβλιοθηκονομία προχωρούσε και η συγκρότηση των καταλόγων εξελισσόταν τόσο η θέση αυτή αποδυναμωνόταν αφού άρχισαν να αναπτύσσονται αρχεία καθιερωμένων αποδόσεων, θεματικά συστήματα κτλ τα οποία μπορούσαν να είναι ανεξάρτητα από τις βιβλιογραφικές εγγραφές. Έτσι, η εγγραφή του πόρου στην ουσία ήταν η σύνθεση επιμέρους εγγραφών. Εντούτοις, εξακολουθούσε να είναι σε ισχύ η αναλογία μίας εγγραφής ανά πόρο αφού στην ουσία η επικεφαλίδα της καθιερωμένης απόδοσης ενσωματωνόταν στη βιβλιογραφική εγγραφή.

Η οριστική και τεκμηριωμένη κατάλυση της λογικής της μίας εγγραφής ανά πόρο εισήχθη με την έλευση των Functional Requirements of Bibliographic Records FRBR¹. Οι FRBR (IFLA, 1998) αντιμετώπισαν τον κατάλογο της βιβλιοθήκης ως ένα δίκτυο μεταξύ αυτόνομων οντοτήτων και όχι ως το άθροισμα αδιαίρετων βιβλιογραφικών εγγραφών. Αυτό σίγουρα αποτελεί την αφετηρία για μεγάλη αλλαγή στη συγκρότηση των καταλόγων. Δεν υπάρχει πλέον μία εγγραφή ανά πόρο αλλά η περιγραφή τού πόρου είναι ένα σύνολο οντοτήτων οι οποίες σχετίζονται μεταξύ τους και τα στοιχεία των οντοτήτων δεν προκύπτουν κατ' ανάγκη από τον συγκεκριμένο πόρο. Δύο ενδεικτικά παραδείγματα, για το ότι οι πληροφορίες δεν προκύπτουν από ένα πόρο, είναι η ημερομηνία δημιουργίας του Έργου και ο ομοιόμορφος τίτλος. Είναι πολύ πιθανό πάνω στο καταλογογραφούμενο τεκμήριο να μην υπάρχει η λέξη «Βίβλος» και όμως αυτή να χρησιμοποιείται ως τίτλος (ομοιόμορφος) για την ευρετηρίαση του τεκμηρίου στον κατάλογο της βιβλιοθήκης. Αντίστοιχα μπορεί να μην αναφέρεται πάνω στο τεκμήριο η ημερομηνία της πρώτης συγγραφής του αλλά, σύμφωνα με τους FRBR, είναι μια ιδιότητα της οντότητας του Έργου. Έτσι, η παρούσα εργασία θεωρεί πως οι FRBR πρέπει να ιδωθούν «όχι ως απλά μεταδεδομένα τα οποία περιγράφουν έναν πόρο αλλά ως ένα είδος γνώσης γύρω από τον πόρο» (Peronakis, 2012, p. 599). Άρα, πλέον, το ζητούμενο είναι να δούμε την καταλογογράφηση ως διαδικασία παρατηρήσεων επί των πόρων (Murray & Tillett, 2011, p. 171) και όχι ως διαδικασία αντιγραφής πληροφοριών από τους πόρους.

Παράλληλα με την ενασχόληση των βιβλιοθηκών με τους FRBR συντελείται σε ένα ευρύτερο πλαίσιο μια ακόμη μεγαλύτερη αλλαγή. Αναπτύσσεται από το World Wide Web Consortium (W3C) το Resource Description Framework (RDF) και αρχίζει ένας διευρυμένος διάλογος γύρω από το σημασιολογικό ιστό (semantic web). Το RDF ως “μοντέλο δεδομένων για μεταδεδομένα” (metadata data model) βασίζεται στην αρχή της τριπλέτας (triple). Δηλαδή, δίνει τη δυνατότητα να συνδεθούν δύο οντότητες μεταξύ τους με μία σχέση. Κάθε τριπλέτα αποτελείται από τη σειρά «υποκείμενο-κατηγορημα-αντικείμενο» (subject-predicate-object) όπου κάθε υποκείμενο ή αντικείμενο μπορεί να αποτελέσει μέρος μιας νέας τριπλέτας. Έτσι παρέχεται η δυνατότητα να δημιουργηθεί ένας γράφος που θα συσχετίζει τους κόμβους/οντότητες που τον αποτελούν.

Η δυνατότητα να δούμε τους καταλόγους των βιβλιοθηκών ως δίκτυο συσχετίσεων μέσω της αρχής του RDF δίνει τη δυνατότητα στις βιβλιοθήκες να αξιοποιήσουν με νέους τρόπους τα δεδομένα που έχουν και να τα εμπλουτίσουν ώστε να είναι σε θέση να παρέχουν πολύ περισσότερες υπηρεσίες και να απαντούν περισσότερα ερωτήματα. Είναι σημαντικό να αξιοποιηθούν οι λογικές του σημασιολογικού ιστού από την υπάρχουσα τεχνογνωσία των βιβλιοθηκών ώστε να ενταχθούν τα δεδομένα τους σε

¹ Σε αντίθεση με τη διεθνή ενασχόληση γύρω από τους FRBR η ελληνική βιβλιογραφία είναι σχετικά περιορισμένη. Μια βασική εισαγωγή μπορεί να βρει κάποιος σε δύο εργασίες με τη συμμετοχή του υπογράφοντος (Πεπονάκης, Στάικος, Σφακάκης, & Καπιδάκης, 2010; Πεπονάκης, 2010). Αντιθέτως η διεθνής βιβλιογραφία γύρω από το θέμα περιλαμβάνει αρκετές εκατοντάδες εργασίες. Για τις σύγχρονες εξελίξεις στην αγγλόφωνη βιβλιογραφία ο αναγνώστης μπορεί να συμβουλευτεί το έργο «The FRBR Family of Conceptual Models: Toward a Linked Bibliographic Future» (Smiraglia, Riva, & Žumer, 2013).

αυτόν και να διαλειτουργήσουν τόσο μεταξύ τους όσο και στο πλαίσιο των φορέων πολιτισμικής κληρονομιάς αλλά και ευρύτερα.

Από τη μία εγγραφή ανά πόρο στα δίκτυα σχέσεων μεταξύ αυτόνομων οντοτήτων

Record based και RDF προσεγγίσεις

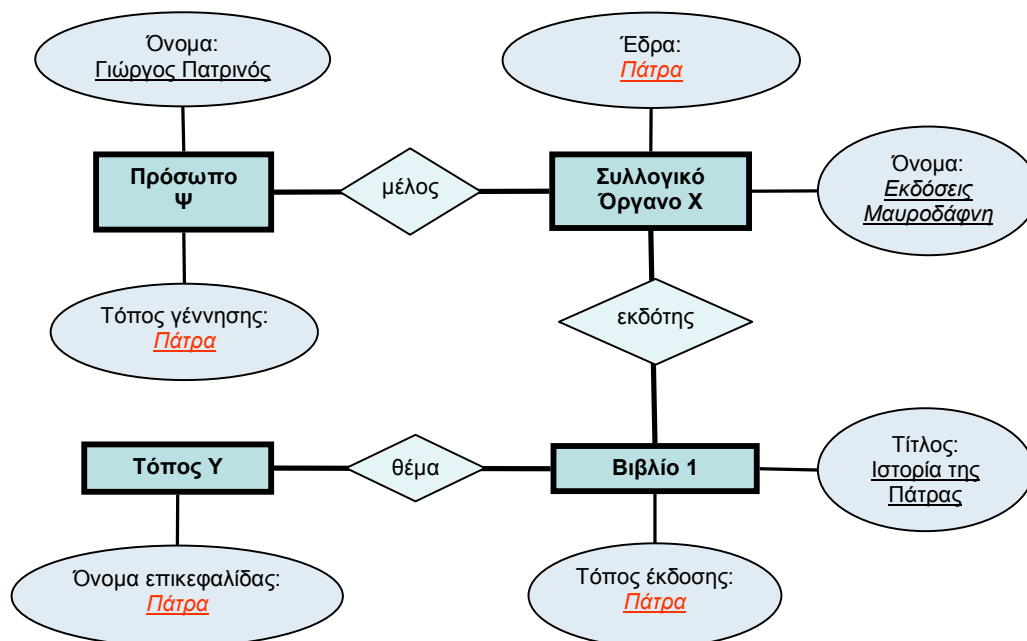
Η διαφορά ανάμεσα στις προσεγγίσεις του Entity Relationship model (ER) και του Object-oriented (OO) είναι αρκετά παλιά (Shoval & Shiran, 1997) και τεκμηριωμένη. Η τρέχουσα υποενότητα στοχεύει κυρίως στην αποσαφήνιση της διάκρισης ανάμεσα στις περιγραφές που βασίζονται σε εγγραφές (record based) και σε εκείνες που ακολουθούν τη λογική της RDF και μάλιστα όπως αυτή εκφράζεται μέσω των Linked Data. Πριν τη συνέχεια είναι αναγκαίο να δηλωθεί σαφώς πως είτε μιλάμε για MARC, DC, MODS ή άλλο αντίστοιχο σχήμα η ουσία είναι κατά βάση η ίδια εφόσον όλες οι προηγούμενες περιπτώσεις αφορούν record based πρακτικές. Όμως η περιγραφή με την RDF έχει τελείως διαφορετική αφετηρία. Όπως αναφέρει ο Baker *“η RDF σχεδιάστηκε για να κάνει δηλώσεις για την πραγματικότητα ενώ το Dublin Core [και τα υπόλοιπα σχήματα μεταδεδομένων] για να καθορίζουν το περιεχόμενο των εγγραφών μεταδεδομένων”* (Baker, 2012, p. 121).

Το ερώτημα που τίθεται είναι: οι FRBR σε ποια από τις δύο προσεγγίσεις ανήκουν; Προφανώς και στην πρώτη αφού ως βάση για την υλοποίησή τους χρησιμοποιήθηκε το ER μοντέλο. Όμως, όπως αναφέρθηκε στην εισαγωγή του παρόντος άρθρου, βασίζονται στη λογική πως η περιγραφή τού πόρου είναι ένα σύνολο οντοτήτων οι οποίες σχετίζονται μεταξύ τους. Άρα, παρότι επέλεξαν το ER, οι εγγραφές που είναι δομημένες με βάση τους FRBR διαθέτουν αρκετά ψήγματα της RDF προσέγγισης σε σημασιολογικό επίπεδο.

Για να καταστεί σαφής η δυσκολία έκφρασης των FRBR στη λογική της RDF παραθέτουμε ένα παράδειγμα χρησιμοποιώντας τον Τόπο όπως αυτός εκφράζεται μέσα από τις entity relationship λογικές των FRBR (IFLA, 1998) και FRAD (IFLA, 2009). Σύμφωνα με τους ορισμούς των FRBR και FRAD ο Τόπος υπάρχει ως οντότητα αλλά και ως ιδιότητα σε αρκετές οντότητες. Συγκεκριμένα Τόπος υπάρχει στο “Τόπος δημοσίευσης / διάθεσης” ως ιδιότητα του *Manifestation*, “Τόπος γέννησης, Τόπος θανάτου, Τόπος διαμονής” ως ιδιότητα του *Προσώπου*, “Τόπος που συνδέεται με την οικογένεια” ως ιδιότητα της *Οικογένειας*, “Τόπος που σχετίζεται με το συλλογικό όργανο” ως ιδιότητα του *Συλλογικού Οργάνου*, “Όρος για τον τόπο” ως ιδιότητα του *Τόπου*, “Τόπος προέλευσης του έργου” ως ιδιότητα του *Έργου*. Αν θεωρήσουμε και το “Τοποθεσία του φορέα = Location of agency” ως τόπο, τότε μετράμε εννιά ιδιότητες οι οποίες μπορεί να αναφέρονται στον ίδιο τόπο, όπως είναι, για παράδειγμα, μια πόλη. Με δεδομένο πως το ER δεν επιτρέπει διασυνδέσεις μεταξύ των ιδιοτήτων δεν είναι δυνατό να δηλωθεί/καταδειχθεί πως πρόκειται για την ίδια πόλη. (Peronakis, 2013). Έτσι είναι αναγκαίο να γίνει έλεγχος της τιμής (δηλαδή του ονόματος) για να

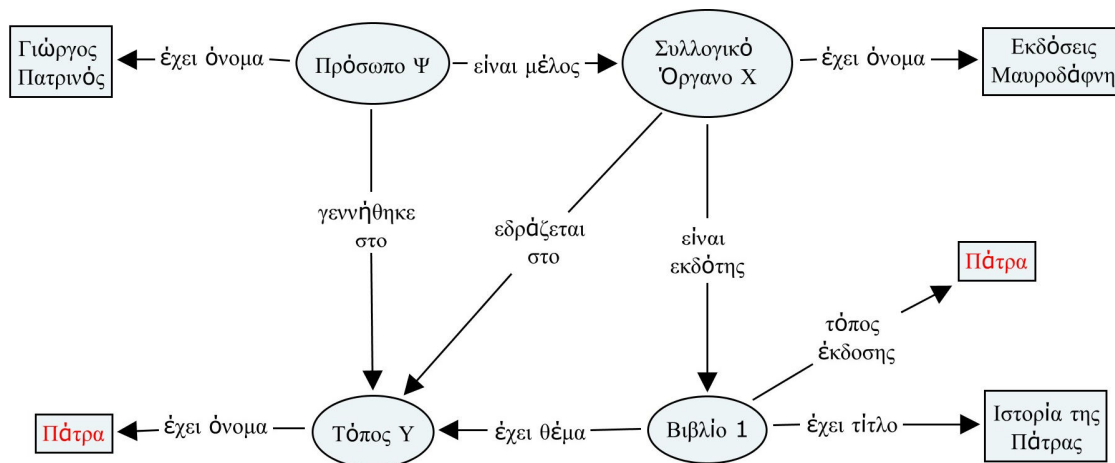
καταλήξουμε πως ίδια τιμή σημαίνει ίδια πόλη. Όμως, όπως είναι γνωστό, η ονομασία δεν είναι επαρκές κριτήριο για τη διάκριση μεταξύ δύο ονομάτων (Doerr, Riva, & Žumer, 2012) και είναι εύκολο να γίνει σύγχυση ανάμεσα στην Αθήνα της Ελλάδας και εκείνης της Γεωργίας (Georgia) των ΗΠΑ.

Με βάση όσα αναφέρθηκαν στην προηγούμενη παράγραφο -ακολουθώντας τις οντότητες, τις ιδιότητες και τις σχέσεις των FRBR και των FRAD- στο σχήμα 1 απεικονίζεται ένα παράδειγμα. Στο παράδειγμα αυτό η πόλη της Πάτρας εμφανίζεται τέσσερις φορές ως ιδιότητα διαφόρων οντοτήτων χωρίς τη δυνατότητα να δηλωθεί ότι πρόκειται για την ίδια πόλη.



Σχήμα 1: Αναπαράσταση ER με βάση τους FRBR και FRAD

Διατηρώντας τη σημασιολογία του σχήματος 1 επανακαθορίζουμε τις οντότητες στο σχήμα 2 εγκαταλείποντας τη λογική του ER και περνάμε σε λογική RDF. Έτσι δίνεται η δυνατότητα να έχουμε φορά (κατεύθυνση) στις σχέσεις και, επί της ουσίας, να προάγουμε κάθε ιδιότητα σε οντότητα. Αυτό μας δίνει τη δυνατότητα να συμπεριφερθούμε στα περισσότερα στοιχεία της περιγραφής ως authorities. Πρέπει όμως όλες οι επαναλήψεις της Πάτρας στο σχήμα 1 να αντιστοιχηθούν στον «Τόπο Υ» του σχήματος 2; Όπως φαίνεται στο σχήμα 2 κάτι τέτοιο δεν έγινε. Οι τέσσερις τιμές της Πάτρας του σχήματος 1 έγιναν δύο. Δεν έγιναν μία διότι ακολουθήθηκε η οδηγία των AACR οι οποίοι υποδεικνύουν ότι η πληροφορία του τόπου έκδοσης του Manifestation πρέπει να γράφεται στο περιγραφικό τμήμα της εγγραφής με τη μορφή που υπάρχει στο τεκμήριο. Εάν οι κανόνες καταλογογράφησης που ακολουθήθηκαν δεν απαιτούσαν την αναγραφή του τόπου έκδοσης και με τη μορφή που αναγράφεται στο τεκμήριο, τότε η Πάτρα θα είχε μία και μόνη επανάληψη.



Σχήμα 2: Αναπαράσταση σε RDF των οντοτήτων του σχήματος 1. Στο σχήμα οι ελλείψεις αναπαριστούν πόρους, οι ευθείες γραμμές ιδιότητες και τα τετράγωνα τιμές.

Από τα παραπάνω καθίσταται σαφές ότι οι FRBR παρότι είναι πραγματικά καινοτόμοι δεν είναι εκφρασμένοι με βάση τις σύγχρονες απαιτήσεις, έτσι η σημασιολογία τους έχει «απώλειες» εξαιτίας των περιορισμών της σύνταξης του επιλεγμένου μοντέλου (ER). Αυτός ακριβώς είναι και ο λόγος που η εναρμόνιση των FRBR με το CIDOC-CRM, η οποία οδήγησε στους FRBRoo (Bekiari, Doerr, & Le Boeuf, 2010; Doerr & LeBoeuf, 2007), είναι πολύ σημαντική συμβολή στο να εκφράσουμε τη σημασιολογία των FRBR σε όρους σημασιολογικού ιστού.

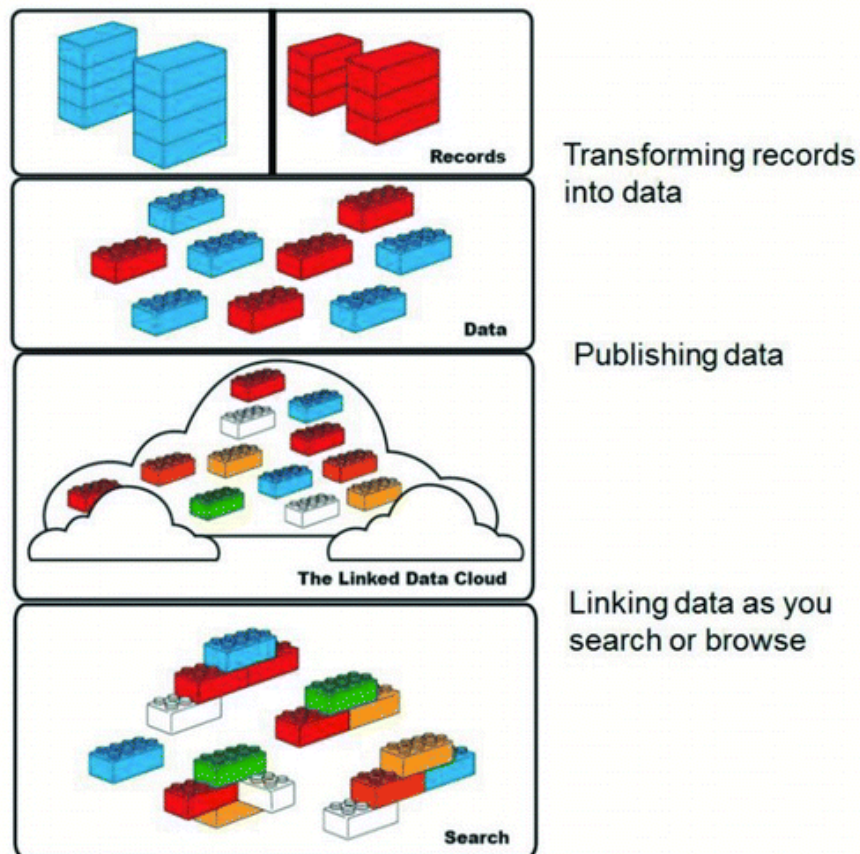
Linked Data: συνδέσεις αντί για download

Στο πλαίσιο του σημασιολογικού ιστού (semantic web) καθένας από τους πόρους-οντότητες του σχήματος 2 μπορεί να χαρακτηριστεί μοναδικά από ένα Uniform Resource Identifier (URI). Όταν το URI αυτό βασίζεται σε http τότε είμαστε ένα βήμα εγγύτερα στα Linked Data (συνδεδεμένα δεδομένα). Τα Linked Data ορίζονται ως ένα “σύνολο βέλτιστων πρακτικών για να δημοσιεύονται και να διασυνδέονται δομημένα δεδομένα στο διαδίκτυο” (Heath & Bizer, 2011).

Μία από τις ουσιαστικές καινοτομίες των Linked Data είναι ότι δεν είναι πλέον απαραίτητο να «κατεβάζουμε» εγγραφές. Μπορούμε απλά να δηλώσουμε, με τη χρήση ενός http URI, ότι αναφερόμαστε σε κάποιον πόρο-οντότητα. Έτσι στην ουσία έχουμε τη δυνατότητα να χρησιμοποιήσουμε το URI εν είδει αντωνυμίας για να αναφερθούμε σε κάτι συγκεκριμένο, χωρίς να χρειάζεται να το περιγράψουμε σε κάθε αναφορά του. Κάθε φορά που είναι αναγκαία η ανάκτηση στοιχείων μπορούμε να ανατρέξουμε στην περιγραφή της συγκεκριμένης οντότητας και να ανακτήσουμε. Προϋπόθεση βέβαια αποτελεί να περιέχει ή, ακριβέστερα, να παραπέμπει η εγγραφή στα στοιχεία αυτά αφού στην ουσία η εγγραφή με τη σειρά της αποτελείται από άλλους συνδέσμους/διασυνδέσεις και ούτω καθεξής. Με αυτό τον τρόπο έχουμε στην ουσία ένα μεγάλο δίκτυο ή, για να χρησιμοποιήσουμε τα λόγια του Tallerås, ένα “γιγάντιο παγκόσμιο γράφο” (Tallerås, 2013). Έτσι αντί να περιγράφει κάθε βιβλιοθήκη

ανεξάρτητα τους πόρους της, μπορεί να χρησιμοποιεί τις περιγραφές άλλων και να τις επεκτείνει με νέα στοιχεία όπου χρειάζεται.

Για να είναι αποτελεσματική η πλοήγηση σε αυτόν τον παγκόσμιο γράφο «ο πόρος πρέπει να αποδομηθεί στις βασικές του οντότητες και η σύνθεση αυτών των οντοτήτων θα επιτρέψει μια καρποφόρα αναπαράσταση του πόρου και των συνδέσεών του με άλλες οντότητες» (Peronakis, 2013). Αυτή η αποδόμηση απεικονίζεται σχηματικά στο σχήμα 3 όπου τα συστατικά στοιχεία καταλήγουν να δημιουργούν νέες συνάψεις ανάλογα με τις απαιτήσεις των χρηστών στο πλαίσιο των αναζητήσεων τους.



Σχήμα 3: Από τις εγγραφές (records) στα δεδομένα (data): the "lego analogy" (Lampert & Southwick, 2013, p. 240)

Ένα σημαντικό θέμα που τίθεται είναι: αφού μπορούμε να «δείχνουμε» σε μια περιγραφή και αυτό να είναι αρκετό, τότε ποιος πρέπει να φτιάξει αυτή την περιγραφή; Ενδεικτικό και ενδιαφέρον παράδειγμα η εγγραφή του Google Scholar για τον Αριστοτέλη βάσει της οποίας ο Αριστοτέλης είναι "Ancient Greek philosopher" με "Verified email at buffalo.edu"². Μπορούμε να βασιστούμε σε αυτή την πληροφορία και να πιστέψουμε ότι έχει εξακριβωθεί πως ο Αριστοτέλης διαθέτει email στο πανεπιστήμιο του Buffalo;!

² Βλέπε <http://scholar.google.com/citations?user=4rO87mQAAAAJ>.

Η επίδραση της μετάβασης στην παραγωγή και τη διάχυση της Πληροφορίας

Στην προηγούμενη ενότητα τεκμηριώθηκε η αλλαγή που συντελείται και κατέστη προφανές ότι μεταβαίνουμε από τις πολλαπλές καταγραφές σε μεμονωμένες βάσεις δεδομένων προς τη μία και μοναδική εγγραφή στην οποία μπορεί ο καθένας να παραπέμπει ή να την επεκτείνει αλλά όχι να κατεβάζει και να τροποποιεί. Είναι αναγκαίο να διευκρινιστεί ότι με τον όρο εγγραφή δεν εννοούμε βιβλιογραφική εγγραφή αλλά την περιγραφή μιας οντότητας. Η οντότητα αυτή μπορεί να είναι ένα Γεγονός, ένα Πρόσωπο, ένας Τόπος, ένα Έργο, ένα Manifestation κτλ και η περιγραφή της δεν είναι παρά ένα ακόμη δίκτυο συσχετίσεων το οποίο αναπόφευκτα καταλήγει σε κάποιες τιμές (βλέπε σχήμα 2). Το κεντρικό όμως ζητούμενο αυτής της προσέγγισης δεν είναι οι τιμές των στοιχείων αλλά οι συνδέσεις μεταξύ τους. Εξ ου και ο όρος συνδεδεμένα δεδομένα (linked data).

Και επανερχόμαστε στο ερώτημα ότι, αφού πλέον οι εγγραφές δεν οικειοποιούνται-τροποποιούνται σε επιμέρους βάσεις δεδομένων αλλά είναι «σημεία» στα οποία μπορεί να «δείξει» κανείς, ποιος θα τις δημιουργήσει ώστε κάποιος άλλος να τις εντοπίσουν και να «χτίσουν» πάνω σε αυτές; Σε κάθε περίπτωση για να «δημιουργήσει» κάποιος μια περιγραφή πρέπει να έχει την τεχνογνωσία αλλά και πρόσβαση στις απαιτούμενες πληροφορίες. Προφανώς σε κάθε τομέα κάποιος φορέας θα παίξουν ηγετικό ρόλο και οι περιγραφές τους θα γίνουν σημεία αναφοράς για άλλους. Ήδη στο χώρο των βιβλιοθηκών το Virtual International Authority File (VIAF) κατέχει εξέχουσα θέση. Το VIAF όμως απλά επανακωδικοποιεί και διαθέτει πληροφορία που του παρέχουν άλλοι, δηλαδή οι βιβλιοθήκες που συμμετέχουν στο εγχείρημα. Στην υποεπένδυση που ακολουθεί θα εξετάσουμε το πώς διαμορφώνεται η πρωτογενής κωδικοποίηση μέσα στις ακαδημαϊκές βιβλιοθήκες.

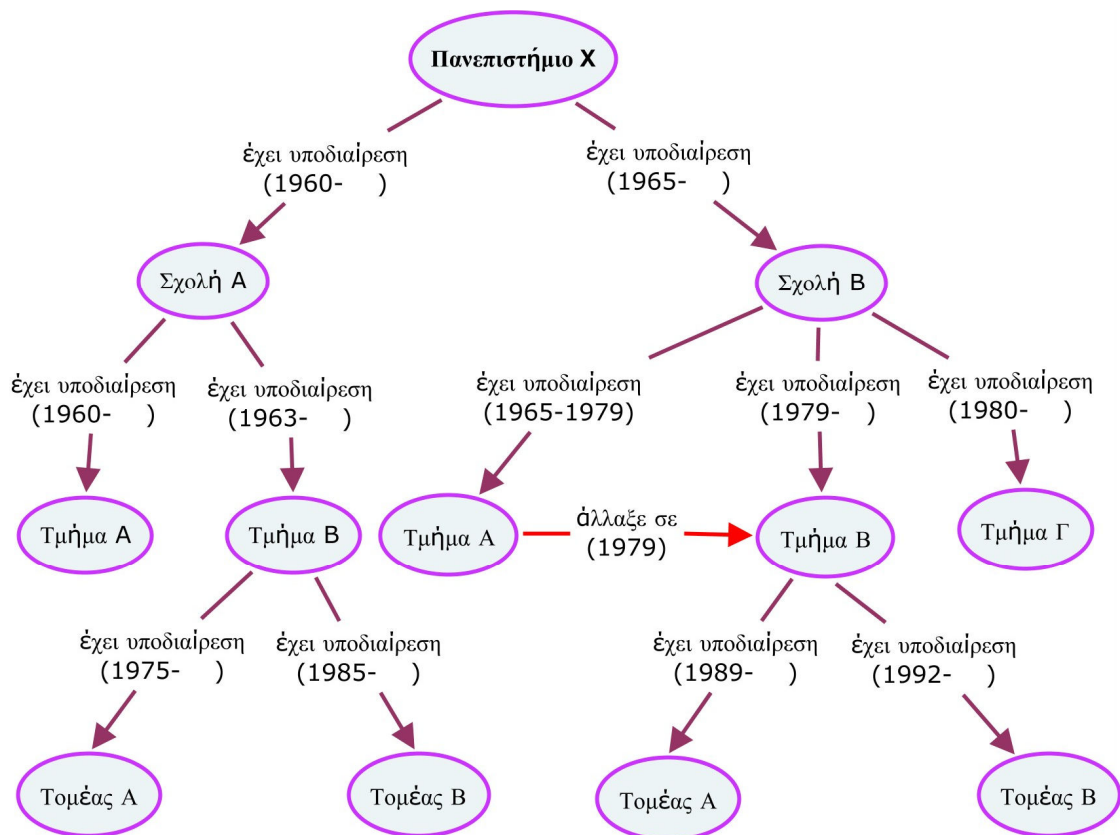
Οι ακαδημαϊκές βιβλιοθήκες ως πάροχοι δεδομένων στο σημασιολογικό ιστό

Στο πλαίσιο της ακαδημαϊκής κοινότητας οι βιβλιοθήκες των ακαδημαϊκών ιδρυμάτων θεωρούνται υπεύθυνες για την καταγραφή μεγάλου μέρους της πληροφορίας που παράγει το οικείο τους ίδρυμα. Οι βασικοί λόγοι για αυτό είναι τρεις. Πρώτον, η βιβλιοθήκη ξέρει πώς να κωδικοποιήσει την πληροφορία, έχει το “know how” της μοντελοποίησης και κωδικοποίησης της πληροφορίας. Δεύτερον, έχει πρόσβαση στην πρωτογενή πληροφορία άρα μπορεί να εγγυηθεί την ακεραιότητά της (integrity). Τρίτον είναι ο φορέας που οι τρίτοι θα εμπιστευθούν ώστε να «χτίσουν» πάνω σε αυτή την πληροφορία (Peronakis, 2013). Άλλωστε η εμπιστοσύνη βρίσκεται στην κορυφή της Πυραμίδας του Σημασιολογικού Ιστού (Semantic Web Stack) σύμφωνα με το σχήμα του μέντορα του ιστού (σημασιολογικού και μη) Tim Berners-Lee (βλέπε <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>).

Οι ακαδημαϊκές βιβλιοθήκες ήδη περιγράφουν τα συγγράμματα που παράγονται από το πανεπιστήμιο (πτυχιακές, μεταπτυχιακές εργασίες, διδακτορικά κτλ) αλλά με τη λογική να καταστήσουν διαθέσιμο το περιεχόμενό τους στην ακαδημαϊκή κοινότητα.

Όμως, πλέον, οι περιγραφές των οντοτήτων έχουν οι ίδιες περιεχόμενο. Έτσι τα μεταδεδομένα τους μετατρέπονται σε δεδομένα τα οποία μπορούν να χρησιμοποιηθούν ως υπηρεσίες προστιθέμενης αξίας και όχι να οδηγούν απλά και μόνο στα τεκμήρια.

Ας πάρουμε ως παράδειγμα την ακαδημαϊκή δομή ενός πανεπιστημίου, δηλαδή την κατάτμησή του σε σχολές, τμήματα κτλ. Με πόση ακρίβεια αναπαριστά τη δομή του πανεπιστημίου και την κατάτμησή του σε σχολές και τμήματα το αρχείο καθιερωμένων αποδόσεων (authority file) της βιβλιοθήκης; Διότι η βιβλιοθήκη του πανεπιστημίου είναι η αρμόδια να παρέχει την εν λόγω δομή εφόσον έχει και την πρόσβαση στις απαιτούμενες πληροφορίες και την τεχνογνωσία της κωδικοποίησης. Το σχήμα 4 απεικονίζει γραφικά μια τέτοια δομή.



Σχήμα 4: Αναπαράσταση μέρους της δομής ενός πανεπιστημίου

Με αυτόν τον τρόπο τα εν λόγω δεδομένα είναι διαθέσιμα ως δομημένη πληροφορία στο σημασιολογικό ιστό και μπορούν να χρησιμοποιηθούν από τρίτους. Η ακαδημαϊκή βιβλιοθήκη μπορεί να εξασφαλίσει την ακεραιότητα (integrity) της πληροφορίας και μέσω της πιστοποιημένης προέλευσης (provenance) της πληροφορίας να αυξηθεί η εμπιστοσύνη (trust) των τρίτων σε αυτήν. Έτσι, για παράδειγμα, η αναπαράσταση της δομής του Πανεπιστημίου από την βιβλιοθήκη του επιτρέπει την αξιοποίηση αυτής της αναπαράστασης με άπειρους τρόπους. Ενδεικτικά παραδείγματα αυτής της χρήσης είναι ο εντοπισμός των μεταπτυχιακών εργασιών ενός τομέα ή μίας

σχολής, η μοντελοποίηση των φοιτητών ανά τμήμα, η κατανομή του γένους των καθηγητών ανά σχολή, οι συγκρίσεις ανάμεσα σε δομές πανεπιστημίων. Αυτό ακριβώς σημαίνει επαναχρησιμοποίηση (reuse) των δεδομένων στο σημασιολογικό ιστό. Εάν η προηγούμενη δομή δεν είναι διαθέσιμη με τρόπο που να μπορεί να χρησιμοποιηθεί άμεσα και ευρέως, καθένας οφείλει να φτιάχνει μια αντίστοιχη δομή κατά περίπτωση (πιθανότατα ελλιπή και μειωμένης αξιοπιστίας) για να την χρησιμοποιήσει. Χαρακτηριστικό παράδειγμα μιας τέτοιας περίπτωσης είναι η πρακτική που εφαρμοζόταν έως τώρα σε μεγάλο βαθμό για τη δημιουργία καθιερωμένων αποδόσεων συλλογικών οργάνων, ιδιαίτερα από φορείς στην ελληνική επικράτεια.

Συμπερασματικά, οι ακαδημαϊκές βιβλιοθήκες είναι εκείνες που οφείλουν να παρουσιάσουν την ακριβή περιγραφή των οντοτήτων που σχετίζονται με τους φορείς που εκπροσωπούν (μέλη ΔΕΠ, μεταπτυχιακές εργασίες κτλ) με αρκετή ακρίβεια και σαφήνεια ώστε να είναι δυνατό οι οντότητες αυτές να διαλειτουργήσουν στο πλαίσιο του σημασιολογικού ιστού.

Συμπεράσματα

Η καταλογογράφηση έχει αλλάξει και ο διαταγμός των βιβλιοθηκών στην ενσωμάτωση των νέων οπτικών μπορεί να έχει μεγάλο τίμημα (Žumer et al., 2011). Η ουσιαστική καινοτομία των FRBR είναι η δυνατότητα να δούμε τους καταλόγους των βιβλιοθηκών ως ένα δίκτυο μεταξύ αυτόνομων οντοτήτων. Από την άλλη, η καινοτομία των Linked Data είναι η δυνατότητα να καθιστούμε σχεδόν οτιδήποτε οντότητα και στη συνέχεια να το διασυνδέουμε, χρησιμοποιώντας διάφορες σχέσεις, με άλλες οντότητες. Οι διασυνδέσεις αυτές δεν λαμβάνουν χώρα στα πλαίσια μιας βάσης δεδομένων ή ενός γνωστικού πεδίου αλλά έχουν καθολική ισχύ. Έτσι, αντί να φτιάχνουμε απομονωμένες βάσεις δεδομένων, μπορεί να δημιουργηθεί ένα παγκόσμιο δίκτυο όπου όλα διαλειτουργούν μεταξύ τους και κάθε τι μπορεί να αποτελέσει βάση για να «χτιστεί» κάτι άλλο πάνω του.

Όμως για να είναι τα δεδομένα των βιβλιοθηκών αρκετά στέρεα ώστε να χρησιμοποιηθούν ευρέως στο πλαίσιο του σημασιολογικού ιστού επιβάλλεται οι βιβλιοθήκες να δημιουργούν ακριβείς και λεπτομερείς περιγραφές.

Εάν θέλουμε “*pidgin*³ για ψηφιακούς τουρίστες” (Baker, 2000) το Dublin Core είναι επαρκές. Αλλά αν θέλουμε να πραγματοποιήσουμε μια βαθύτερη ανάλυση τα “*pidgins*” δεν είναι κατάλληλα για επιστημονικό λόγο. Έτσι πρέπει να δημιουργήσουμε πιο τυπικές (formal) προσεγγίσεις -αυτή την εποχή εκφρασμένες σε όρους τυπικών οντολογιών- οι οποίες θα μας επιτρέψουν διαφορετικά επίπεδα αφαίρεσης. Με αυτό τον τρόπο είναι δυνατό να γίνουμε πιο συγκεκριμένοι όταν, για παράδειγμα, εστιάζουμε στα διδακτορικά απ’ ότι όταν αναφερόμαστε στις μεταπτυχιακές σπουδές

³ «Pidgin» είναι η απλοποιημένη γλώσσα η οποία χρησιμοποιείται για την επικοινωνία μεταξύ δύο ή περισσότερων ομάδων οι οποίες δεν διαθέτουν κάποια κοινή γλώσσα. Η ετυμολογία της λέξης δεν είναι βέβαιη αλλά η πιο αποδεκτή ερμηνεία αναφέρει πως προέρχεται από την αγγλική λέξη «business» αλλοιωμένη από την κινέζικη προφορά. Βλέπε <http://en.wikipedia.org/wiki/Pidgin>.

γενικότερα κοκ. Στο κορυφαίο επίπεδο της αφαίρεσης μπορεί να υπάρχει ένα γενικό απλό σχήμα (όπως το DC) όπου η πλούσια σημασιολογία των αρχικών σχημάτων αντιστοιχίζεται στα γενικά στοιχεία κατόπιν πολλαπλών απλοποιήσεων. Συνεπώς αυτά τα επίπεδα θα επιτρέψουν στους φορμαλισμούς μας να εκτείνονται μεταξύ της ακριβούς πολυπλοκότητας και των ασαφών απλοποιήσεων (ridgin) ανάλογα με τις δεδομένες ανάγκες. Εάν τα ελεγχόμενα λεξιλόγια μας αποτελούνται μόνο από ridgin τότε είναι πολύ δύσκολο να τα αναβαθμίσουμε σε επιστημονικό λόγο. Από την άλλη εάν δεν έχουμε επίπεδα αφαίρεσης δεν είναι εύκολο να απλοποιήσουμε τον επιστημονικό λόγο με σκοπό να επικοινωνήσει με άλλους χώρους και κοινότητες (Peronakis, 2013).

Μέσα από την παραπάνω προσέγγιση τα δεδομένα των βιβλιοθηκών (ακαδημαϊκών και μη) μπορούν να συμβάλουν καθοριστικά στην ανάπτυξη του σημασιολογικού ιστού και να αξιοποιηθούν ως βάση για την παροχή πλήθους υπηρεσιών προστιθέμενης αξίας.

Βιβλιογραφία

- Baker, T. (2000). A Grammar of Dublin Core. *D-Lib Magazine*, 6(10). doi:10.1045/october2000-baker
- Baker, T. (2012). Libraries, Languages of Description, and Linked Data: a Dublin Core Perspective. *Library Hi Tech*, 30(1), 116–133. doi:10.1108/07378831211213256
- Bekiari, C., Doerr, M., & Le Boeuf, P. (2010). *FRBR Object-Oriented Definition and Mapping FRBRER* (version 1.0.1) (p. 153). International Working Group on FRBR and CIDOC CRM Harmonisation. Retrieved from http://www.cidoc-crm.org/docs/frbr_oo/frbr_docs/FRBRoo_V1.0.1.pdf
- Doerr, M., & LeBoeuf, P. (2007). Modelling Intellectual Processes: The FRBR - CRM Harmonization. In C. Thanos, F. Borri, & L. Candela (Eds.), *Digital Libraries: Research and Development* (pp. 114–123). Retrieved from http://dx.doi.org/10.1007/978-3-540-77088-6_11
- Doerr, M., Riva, P., & Žumer, M. (2012). FRBR Entities: Identity and Identification. *Cataloging & Classification Quarterly*, 50(5-7), 517–541. doi:10.1080/01639374.2012.681252
- Heath, T., & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. San Rafael: Morgan & Claypool. Retrieved from <http://dx.doi.org/10.2200/S00334ED1V01Y201102WBE001>
- IFLA. (1998). *Functional Requirements for Bibliographic Records: Final Report* (As amended and corrected through February 2009., Vol. 19). München: K.G. Saur.
- IFLA. (2009). *Functional Requirements for Authority Data: A Conceptual Model*. (G. E. Patton, Ed.). München: K.G. Saur.

- Lampert, C. K., & Southwick, S. B. (2013). Leading to Linking: Introducing Linked Data to Academic Library Digital Collections. *Journal of Library Metadata*, 13(2-3), 230–253. doi:10.1080/19386389.2013.826095
- Murray, R. J., & Tillett, B. B. (2011). Cataloging Theory in Search of Graph Theory and Other Ivory Towers. *Information Technology and Libraries*, 30(4), 170–184. doi:10.6017/ital.v30i4.1868
- Peponakis, M. (2012). Conceptualizations of the Cataloging Object: A Critique on Current Perceptions of FRBR Group 1 Entities. *Cataloging & Classification Quarterly*, 50(5-7), 587–602. doi:10.1080/01639374.2012.681275
- Peponakis, M. (2013). Libraries' Metadata as Data in the Era of the Semantic Web: Modeling a Repository of Master Theses and PhD Dissertations for the Web of Data. *Journal of Library Metadata*, 13(4), [υπό έκδοση]. doi:10.1080/19386389.2013.846618
- Shoval, P., & Shiran, S. (1997). Entity-relationship and object-oriented data modeling: An experimental comparison of design quality. *Data & Knowledge Engineering*, 21(3), 297–315. doi:10.1016/S0169-023X(97)88935-5
- Smiraglia, R. P. (2009). Bibliocentrism, Cultural Warrant, and the Ethics of Resource Description: A Case Study. *Cataloging & Classification Quarterly*, 47(7), 671–686. doi:10.1080/01639370903112013
- Smiraglia, R. P., Riva, P., & Žumer, M. (Eds.). (2013). *The FRBR Family of Conceptual Models: Toward a Linked Bibliographic Future*. London: Routledge.
- Tallerås, K. (2013). From many records to one graph: heterogeneity conflicts in the linked data restructuring cycle. *Information Research*, 18(3). Retrieved from <http://informationr.net/ir/18-3/colis/paperC18.html>
- Žumer, M., Pisanski, J., Vilar, P., Harej, V., Merčun, T., & Švab, K. (2011). Breaking barriers between old practices and new demands: the price of hesitation. Presented at the 77th IFLA General Conference and Assembly, Puerto Rico. Retrieved from <http://conference.ifla.org/past/2011/80-zumer-en.pdf>
- Πεπονάκης, Μανόλης. (2010). *Σύνθεση FRBR εγγραφών αξιοποιώντας υπάρχουσες βιβλιογραφικές εγγραφές (FRBRization): ομαδοποίηση σχετικών εγγραφών (clustering) και εμφάνισή τους σε on line συστήματα* (Thesis). Ιόνιο Πανεπιστήμιο. Retrieved from <http://eprints.rclis.org/16674/>
- Πεπονάκης, Μανόλης, Στάικος, Παναγιώτης, Σφακάκης, Μιχάλης, & Καπιδάκης, Σαράντος. (2010). FRBRization: αναζητώντας το 'κλειδί' για τον εντοπισμό των Έργων. In *19ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών: Επιστημονικές κοινότητες και βιβλιοθήκες στον κόσμο της κοινωνικής δικτύωσης και συνέργειας* (pp. 221–235). Αθήνα: Βιβλιοθήκη-Υπηρεσία Πληροφόρησης Παντείου Πανεπιστημίου. Retrieved from <http://eprints.rclis.org/16672/>