



## Συνδεδεμένα Δεδομένα: Μια ευκαιρία για τις ελληνικές Βιβλιοθήκες

Δρ. Δημήτριος-Εμμανουήλ Σπανός<sup>α</sup>

<sup>α</sup>Ερευνητικός Συνεργάτης, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών,  
Εθνικό Μετσόβιο Πολυτεχνείο, [dspanos@cn.ntua.gr](mailto:dspanos@cn.ntua.gr)

### Περίληψη

Τα Συνδεδεμένα Δεδομένα αποτελούν μια τεχνολογία που, ενώ εμφανίστηκε πριν μερικά μόνο χρόνια στο χώρο του Παγκόσμιου Ιστού, σήμερα κερδίζει συνεχώς έδαφος, κυρίως σε τομείς όπως η ηλεκτρονική διακυβέρνηση, η οικονομία, η βιοϊατρική και τα μέσα ενημέρωσης, γεγονός που αποδεικνύεται από την ενίσχυση -παγκοσμίως- της τάσης για αξιοποίηση των ευκαιριών που αυτά προσφέρουν. Στο παρόν άρθρο, μετά από μια εισαγωγή στο χώρο των Συνδεδεμένων Δεδομένων και των σχετικών τεχνολογιών, εξετάζονται τα οφέλη που προκύπτουν από την υιοθέτηση των συγκεκριμένων τεχνολογιών από τις Βιβλιοθήκες, όπου η εφαρμογή τους βρίσκει πρόσφορο έδαφος, και προτείνονται μέθοδοι για την υιοθέτηση της συγκεκριμένης ελπιδοφόρας νέας τεχνολογίας.

© 2015 Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών

**Λέξεις-κλειδιά:** συνδεδεμένα δεδομένα, ψηφιακές βιβλιοθήκες, RDF, ιστός δεδομένων, σημασιολογικός ιστός

doi:

### Abstract

*Linked Data represents a novel technology that has emerged in the World Wide Web landscape during the last few years and is already mentioned and applied in diverse domains, such as electronic government, the financial sector, biomedicine and mass media. The application of Linked Data can be extremely advantageous for the library domain and there is already a related global trend in seizing this opportunity. This paper attempts an introduction to the Linked Data movement and related technologies, investigates the benefits arising from the application of such technologies in the library domain and proposes some necessary steps for the adoption of this new promising technology.*

© 2015 Hellenic Academic Libraries Link

**Keywords:** linked data, digital libraries, RDF, web of data, semantic web

## 1. Εισαγωγή

Ένας από τους πρωταρχικούς λόγους δημιουργίας του Διαδικτύου ήταν ο διαμοιρασμός πληροφορίας μεταξύ απομακρυσμένων συστημάτων. Ο Παγκόσμιος Ιστός, όπως τον γνωρίζουμε έως σήμερα, έχει δομηθεί σε πρωτόκολλα και τεχνολογίες Διαδικτύου, όπως το πρωτόκολλο HTTP που προδιαγράφει την επικοινωνία μεταξύ καταναμημένων συστημάτων.

Βασική μονάδα του Παγκόσμιου Ιστού θεωρείται η ιστοσελίδα, η οποία κατά κύριο λόγο προορίζεται για ανάγνωση και επεξεργασία από τους χρήστες. Υπ' αυτό το πρίσμα, ο Παγκόσμιος Ιστός είναι απλώς ένα δίκτυο εγγράφων το οποίο δημιουργείται από τους συνδέσμους που ενώνουν τα έγγραφα-ιστοσελίδες μεταξύ τους. Για την αναζήτηση και εύρεση συγκεκριμένης πληροφορίας ο χρήστης πρέπει να πλοηγηθεί από έγγραφο σε έγγραφο ακολουθώντας τους συνδέσμους, να εξαγάγει από τα έγγραφα τα απαραίτητα δεδομένα και να τα συνδυάσει χειροκίνητα με δεδομένα από άλλα έγγραφα. Τα τελευταία χρόνια η διαδικασία έχει απλοποιηθεί εν μέρει, λόγω της αυξανόμενης διαθεσιμότητας προγραμματιστικών διεπαφών που επιτρέπουν την ανάκτηση δεδομένων σε μορφή εύκολα επεξεργάσιμη μέσω γλωσσών προγραμματισμού, ευνοώντας την ανάπτυξη καινοτόμων εφαρμογών που συνδυάζουν δεδομένα διαφορετικής προέλευσης. Ωστόσο, ο τρόπος λειτουργίας του Παγκόσμιου Ιστού ως ενός ιστού εγγράφων (web of documents) θέτει όρια στην αξιοποίηση του δυναμικού και των ευκαιριών που προσφέρει ο τεράστιος ανεπεξέργαστος όγκος δεδομένων που είναι διαθέσιμα στο Διαδίκτυο.

Αυτός είναι και ο λόγος που ο εμπνευστής του Παγκόσμιου Ιστού, Tim Berners-Lee, προωθεί εδώ και περισσότερο από μία δεκαετία (Berners-Lee, Hendler & Lassila, 2001) την ιδέα ενός ιστού δεδομένων (web of data) ή Σημασιολογικού Ιστού (Semantic Web), όπου πλέον το έγγραφο παύει να συνιστά τη βασική μονάδα. Στο Σημασιολογικό Ιστό το έγγραφο αντικαθίσταται από την έννοια ή οντότητα, που μπορούν να περιγραφούν με δομημένο τρόπο και να συνδεθούν με άλλες οντότητες, συγκροτώντας με αυτόν τον τρόπο ένα γιγάντιο παγκόσμιο δίκτυο ή γράφο εννοιών και δεδομένων. Δηλαδή, ο Σημασιολογικός Ιστός θεμελιώνεται στις ίδιες τεχνολογίες και πρωτόκολλα που χρησιμοποιεί και ο Παγκόσμιος Ιστός, και, επιπρόσθετα, μέσω νέων προτύπων, επιτρέπει την περιγραφή και διαχείριση δεδομένων που δομούνται σε μορφή γράφου.

Στο παρόν άρθρο θα παρουσιαστούν τα Συνδεδεμένα Δεδομένα (Linked Data), δηλαδή *δεδομένα δημιουργημένα στον Ιστό σύμφωνα με συγκεκριμένες πρακτικές, ώστε να μπορούν να συνδεθούν μεταξύ τους και να δημιουργήσουν ένα δίκτυο δεδομένων*, καθώς και τα πρότυπα και οι τεχνολογίες στις οποίες αυτά βασίζονται (Ενότητα 2). Επίσης, θα αναλυθούν οι ωφέλειες της υιοθέτησής τους στο χώρο των Βιβλιοθηκών (Ενότητα 3) και θα προταθούν μέθοδοι παραγωγής Συνδεδεμένων Δεδομένων από υπάρχοντα δεδομένα (Ενότητα 4).

## 2. Συνδεδεμένα δεδομένα

Η τρέχουσα ενότητα αποτελεί μια εισαγωγή στα θεμελιώδη πρότυπα του Σημασιολογικού Ιστού και στην έννοια των Συνδεδεμένων Δεδομένων. Όπως αναφέρθηκε και στην ενότητα 1, η βασική δομική μονάδα του Παγκόσμιου Ιστού είναι το *έγγραφο* (document). Οι πρωταρχικές τεχνολογίες στις οποίες δομείται ο Παγκόσμιος Ιστός είναι το πρωτόκολλο HTTP (HyperText Transfer Protocol), ένα πρότυπο επικοινωνίας για τη μεταφορά εγγράφων στον Ιστό, καθώς και η γλώσσα HTML (HyperText Markup Language), η οποία καθορίζει τη δομή, το περιεχόμενο και την παρουσίαση ενός εγγράφου. Ένα έγγραφο HTML μπορεί να περιέχει *συνδέσμους* (links) προς άλλα έγγραφα, τα οποία αναγνωρίζονται μέσω του URL (Uniform Resource Locator) τους, το οποίο καθορίζει την τοποθεσία τους στον Ιστό. Έτσι, ο Παγκόσμιος Ιστός μπορεί να ιδωθεί ως ένας ιστός ή ένα δίκτυο από έγγραφα HTML. Μάλιστα, οι πρώτες εκδόσεις σε HTML δεν παρείχαν ικανοποιητικούς μηχανισμούς για τη δομημένη περιγραφή των περιεχομένων ενός εγγράφου και τη σύνδεση αυτών μεταξύ τους, αλλά οδηγίες κυρίως για την οπτική αναπαράσταση του εγγράφου. Έτσι, ο χρήστης, είτε για να αναζητήσει μια συγκεκριμένη πληροφορία (π.χ. το ωράριο ενός καταστήματος) είτε για να πραγματοποιήσει μια συγκριτική μελέτη, οφείλει πρώτα να πλοηγηθεί στα κατάλληλα έγγραφα και να τα διαβάσει, ώστε να εντοπίσει τις απαραίτητες πληροφορίες, και, εν συνεχεία μόνος του να επεξεργαστεί τις εν λόγω πληροφορίες ή/και να τις συνδυάσει χειροκίνητα.

Ο εμπνευστής του Παγκόσμιου Ιστού, Sir Tim Berners-Lee, αναγνωρίζοντας τις εγγενείς αδυναμίες του Ιστού, οι οποίες δυσχεραίνουν την επεξεργασία από προγραμματιστικές διαδικασίες της πληροφορίας που υπάρχει σε μορφή HTML εγγράφων, καθώς και την αυτοματοποίηση συνηθισμένων λειτουργιών τις οποίες αναγκάζονται να πραγματοποιούν χειροκίνητα οι χρήστες, συνέλαβε την έννοια ενός Σημασιολογικού Ιστού (Semantic Web) ή ενός ιστού δεδομένων (web of data). Στο Σημασιολογικό Ιστό παρέχεται η δυνατότητα αυστηρής, δομημένης περιγραφής και σύνδεσης, όχι πλέον σε επίπεδο εγγράφου, αλλά σε επίπεδο εννοιών και οντοτήτων. Κατ' αναλογία με την έννοια του URL, κεντρικό ρόλο στον ιστό δεδομένων διαδραματίζει η έννοια του URI (Uniform Resource Identifier), ενός μοναδικού αναγνωριστικού για οτιδήποτε (αντικείμενο, οντότητα ή έννοια) περιγράφεται στο Σημασιολογικό Ιστό, και συνιστά έννοια ευρύτερη του URL.

Στην πορεία υλοποίησης του οράματος του Σημασιολογικού Ιστού εμφανίστηκε το κίνημα των Συνδεδεμένων Δεδομένων (Linked Data) (Heath & Bizer, 2011), το οποίο απέκτησε αρκετή φήμη, τόσο ώστε πλέον θεωρείται συνώνυμο του οράματος για μετάβαση σε ένα γνήσιο Σημασιολογικό Ιστό. Ο όρος Συνδεδεμένα

Δεδομένα (Linked Data) περιλαμβάνει ένα σύνολο βέλτιστων πρακτικών για τη δημοσίευση και διασύνδεση δομημένων δεδομένων στον Παγκόσμιο Ιστό.

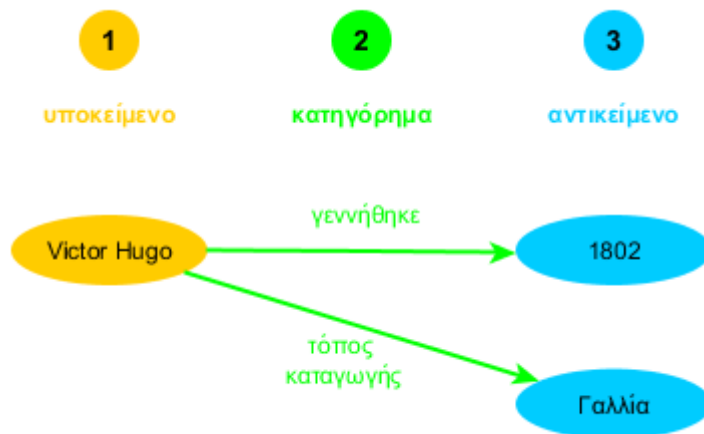
Οι τέσσερις (4) γενικές αρχές δημοσίευσης Συνδεδεμένων Δεδομένων είναι οι ακόλουθες (Berners-Lee, 2006):

1. Η χρήση URIs για την ονομασία οντοτήτων και την αναφορά σε αυτές (όπως ένα URL χρησιμοποιείται ως αναγνωριστικό ενός εγγράφου HTML)
2. Η χρήση URIs που βασίζονται στο πρωτόκολλο HTTP, ώστε αυτά να είναι προσβάσιμα μέσω γνωστών εργαλείων του Παγκόσμιου Ιστού
3. Κατά την προσπέλαση ενός URI, παροχή πρόσθετης πληροφορίας για την οντότητα στην οποία αυτό αναφέρεται
4. Ορισμός συνδέσμων προς άλλες οντότητες, οι οποίες επίσης αναγνωρίζονται με αντίστοιχα URIs

Κύρια συστατικά για την υλοποίηση αυτών των τεσσάρων (4) αρχών είναι το πρωτόκολλο HTTP, ως ενιαίος παγκόσμιος μηχανισμός πρόσβασης, καθώς και νέες τεχνολογίες, όπως το μοντέλο RDF για τη δομημένη περιγραφή οντοτήτων και η γλώσσα SPARQL για την αναζήτηση αυτών στον ιστό δεδομένων.

Το μοντέλο RDF (Resource Description Framework) είναι ένα μοντέλο περιγραφής δομημένης πληροφορίας το οποίο ακολουθεί τη δομή ενός γράφου και βασίζεται στη χρήση URIs για τη μονοσήμαντη αναφορά σε οντότητες. Με βάση τα εν λόγω χαρακτηριστικά του, το RDF ταιριάζει απόλυτα στη λογική του Σημασιολογικού Ιστού και των Συνδεδεμένων Δεδομένων, για το σκοπό αυτό άλλωστε αποτελεί το βασικό μοντέλο για την περιγραφή και αναπαράσταση δεδομένων στο πλαίσιο του υπό ανάπτυξη ιστού δεδομένων. Το κύριο δομικό στοιχείο του RDF μοντέλου είναι η πρόταση (statement) ή τριάδα (triple)· εκφράζει μια συσχέτιση μεταξύ δύο πόρων, από τους οποίους ο πρώτος ονομάζεται *υποκείμενο* και ο δεύτερος αντικείμενο, ενώ το τρίτο συστατικό στοιχείο μιας πρότασης, το *κατηγορημα*, περιγράφει τη φύση της συσχέτισης. Μια RDF πρόταση θα μπορούσε, για παράδειγμα, να εκφράζει το γεγονός ότι ο συγγραφέας Victor Hugo γεννήθηκε το έτος 1802.

Ένα δεύτερο χαρακτηριστικό του RDF μοντέλου, που ήδη αναφέρθηκε, είναι η χρήση URIs ως αναγνωριστικών για τις οντότητες που περιγράφονται σε αυτό. Συνεπώς, στο προηγούμενο παράδειγμα RDF πρότασης, πρέπει να χρησιμοποιηθούν ξεχωριστά URIs που να περιγράφουν την οντότητα Victor Hugo, τη συσχέτιση «γεννήθηκε» και ενδεχομένως και το έτος 1802, προκειμένου να είναι σαφές ότι ο εν λόγω αριθμός αναφέρεται σε κάποια χρονολογία και δεν έχει άλλη σημασία. Άμεση συνέπεια της χρήσης URIs για το μονοσήμαντο προσδιορισμό εννοιών και οντοτήτων είναι το γεγονός ότι μια δεύτερη RDF πρόταση που αναφέρεται στην οντότητα Victor Hugo μπορεί να επαναχρησιμοποιήσει το URI της πρώτης. Με αυτόν τον τρόπο, μεμονωμένες RDF προτάσεις συνδέονται μεταξύ τους δημιουργώντας ένα γράφο, όπως φαίνεται στην Εικόνα 1.



Εικόνα 1. Παράδειγμα RDF γράφου σε αφηρημένη μορφή

Είναι προφανές ότι ο παραπάνω γράφος μπορεί να επεκταθεί, συνδέοντας RDF προτάσεις που έχουν ως υποκείμενο ή αντικείμενό τους είτε τον κόμβο του έτους 1802 είτε τον κόμβο της Γαλλίας. Μια τέτοια πρόταση θα μπορούσε να αναφέρει ότι η Γαλλία (υποκείμενο) αποτελεί μέρος (κατηγορημα) της Ευρώπης (αντικείμενο). Στο πλαίσιο του τρέχοντος άρθρου, κρίνεται άσκοπη μια πλήρης περιγραφή των χαρακτηριστικών του RDF μοντέλου, όπως και η εμβάθυνση σε τεχνικές λεπτομέρειες. Αρκεί μόνο να σημειωθεί ότι RDF γράφοι μπορούν να

κωδικοποιηθούν σε διάφορες μηχαναγνώσιμες μορφές (ενδεικτικά παραδείγματα αποτελούν οι Turtle, RDFa, RDF/XML), προσφέροντας ευελιξία στην προγραμματιστική επεξεργασία τους. Στο υπόλοιπο του άρθρου, θα αναπαριστώνται RDF τριάδες και γράφοι σύμφωνα με τη σύνταξη Turtle, η οποία είναι πιο συνοπτική από τη γραφική αναπαράσταση της Εικόνας 1. Ενδεικτικά, αγνοώντας προς το παρόν την εύρεση κατάλληλων URIs για καθένα από τα συστατικά στοιχεία του γράφου της Εικόνας 1, ο εν λόγω γράφος είναι ισοδύναμος με τις ακόλουθες δύο (2) τριάδες:

```
<Victor_Hugo> <γεννήθηκε> <1802>;
    <έχει καταγωγή> <Γαλλία>.
```

Εντούτοις, ένας από τους καθοριστικούς παράγοντες για την παραγωγή Συνδεδεμένων Δεδομένων υψηλής ποιότητας είναι η επιλογή και χρήση των URIs που αντιστοιχούν σε καθένα από τα συστατικά μιας RDF τριάδας. Τα πλεονεκτήματα της επαναχρησιμοποίησης δημοφιλών URIs για την αναφορά σε οντότητες και συσχετίσεις είναι προφανή: αφενός επιτρέπουν τη σύνδεση απομονωμένων RDF γράφων και αφετέρου διευκολύνουν τη σημασιολογική διαλειτουργικότητα γράφων που έχουν προέλθει από διαφορετικές πηγές. Μέχρι στιγμής έχουν γίνει σημαντικές προσπάθειες δημιουργίας Συνδεδεμένων Δεδομένων, με την ανάπτυξη αρκετών λεξιλογίων που καλύπτουν πολλές Θεματικές Περιοχές, μεταξύ άλλων και τις Βιβλιοθήκες. Ειδικότερα για το βιβλιοθηκονομικό τομέα ενδιαφέροντος, τα εν λόγω λεξιλόγια μπορούν να διακριθούν σε *λεξιλόγια τιμών* (value vocabularies) και σε *σύνολα στοιχείων μεταδεδομένων* (metadata element sets) (Baker et al., 2011).

Τα λεξιλόγια τιμών αποτελούν σύνολα δεδομένων που ορίζουν οντότητες και έννοιες που χρησιμοποιούνται σε βιβλιοθηκονομικές εγγραφές, όπως πρόσωπα, θέματα και τοποθεσίες. Λεξιλόγια τιμών θεωρούνται μεταξύ άλλων οι θησαυροί και οι ταξινομίες όρων, καθώς και οι κατάλογοι καθιερωμένων όρων. Γνωστά παραδείγματα λεξιλογίων διαθέσιμων σε μορφή Συνδεδεμένων Δεδομένων αποτελούν οι θεματικοί όροι της Βιβλιοθήκης του Κογκρέσου (LCSH), το δεκαδικό σύστημα ταξινόμησης Dewey (DDC), το διεθνές αρχείο καθιερωμένων όρων VIAF και η γεωγραφική βάση δεδομένων GeoNames.

Ενώ τα λεξιλόγια τιμών αποτελούν, κατά κάποιον τρόπο, ευρετήρια URIs που μπορούν να χρησιμοποιηθούν είτε στη θέση υποκειμένου είτε αντικειμένου μιας RDF τριάδας, τα σύνολα στοιχείων μεταδεδομένων είναι σύνολα κλάσεων και ιδιοτήτων· μάλιστα οι ιδιότητες είναι αυτές που χρησιμοποιούνται στη θέση του κατηγορήματος μιας RDF τριάδας. Στον ευρύτερο χώρο του Σημασιολογικού Ιστού τα σύνολα στοιχείων μεταδεδομένων είναι γνωστά και ως *οντολογίες* (ontologies). Μια οντολογία αποτελεί την αυστηρή περιγραφή μιας Θεματικής Περιοχής και, εκτός από σύνολα κλάσεων και ιδιοτήτων, μπορεί επίσης να περιέχει και ταξινομίες κλάσεων και ιδιοτήτων, καθώς και αξιώματα που πρέπει να ικανοποιούν οι εν λόγω κλάσεις και ιδιότητες. Οι οντολογίες μπορούν να εκφραστούν σε γλώσσες αναπαράστασης γνώσης, όπως η RDFS (Brickley & Guha, 2014) και η OWL (Hitzler, Krötzsch, Parsia, Patel-Schneider, & Rudolph, 2012), γλώσσες διαφορετικής εκφραστικότητας ωστόσο. Η χρήση όρων οντολογιών σε RDF γράφους καθιστά δυνατή την εφαρμογή συλλογισμού (reasoning) για την εξαγωγή νέων RDF προτάσεων που συνάγονται από τις ρητά δηλωθείσες προτάσεις και για τον έλεγχο της συνέπειας ενός RDF γράφου με βάση τα αξιώματα των εμπλεκόμενων οντολογιών. Χαρακτηριστικά παραδείγματα οντολογιών που χρησιμοποιούνται από τις Βιβλιοθήκες είναι τα εξής: Dublin Core, SKOS (Simple Knowledge Organization System), BIBO (Bibliographic Ontology) και FOAF (Friend of a Friend).

Η γνώση των υπάρχοντων διαθέσιμων λεξιλογίων είναι απαραίτητη για τους επιστήμονες πληροφόρησης ή τους μηχανικούς γνώσης - που καλούνται να μετατρέψουν υπάρχοντα σύνολα δεδομένων σε RDF γράφους, συντελώντας στη δημιουργία πραγματικά Συνδεδεμένων Δεδομένων υψηλής ποιότητας, τα οποία μπορούν άμεσα να ενσωματωθούν στο παγκόσμιο Σύννεφο Ανοικτών Συνδεδεμένων Δεδομένων (Linked Open Data Cloud). Έως σήμερα καταγράφονται αρκετά παραδείγματα ευρωπαϊκών Βιβλιοθηκών που έχουν υιοθετήσει τη φιλοσοφία των Συνδεδεμένων Δεδομένων, αποτελώντας παράδειγμα και οδηγό για τις ελληνικές Βιβλιοθήκες. Άξια αναφοράς είναι τα παραδείγματα της Εθνικής Βιβλιοθήκης της Σουηδίας (Malmsten, 2008), της Γερμανικής Εθνικής Βιβλιοθήκης (<http://www.dnb.de/EN/lds>), της Εθνικής Βιβλιοθήκης της Ισπανίας (Vila-Suero & Gomez-Perez, 2013), καθώς και της Βρετανικής Βιβλιοθήκης (<http://bnb.data.bl.uk/>), οι οποίες δημοσιεύουν τους βιβλιογραφικούς καταλόγους τους και τα αρχεία καθιερωμένων όρων τους - θεματικούς όρους και πρόσωπα - σε μορφή Συνδεδεμένων Δεδομένων, υιοθετώντας σε γενικές γραμμές τη διαδικασία που σκιαγραφείται στην Ενότητα 4.

### 3. Οφέλη για Βιβλιοθήκες

Η τεχνολογία των Συνδεδεμένων Δεδομένων εμφανίζει από τη φύση της ποικίλα πλεονεκτήματα, τόσο για τον ιδιοκτήτη δεδομένων που αποφασίζει να τα δημοσιεύσει σε μορφή Συνδεδεμένων Δεδομένων όσο και για τον τελικό

χρήστη, ο οποίος τα καταναλώνει και τα εκμεταλλεύεται. Παραδείγματος χάριν, ο πρώτος μπορεί να επιτύχει με μικρό κόστος ολοκλήρωση και σύνδεση των δεδομένων του με εξωτερικά ετερογενή -δομημένα, ημιδομημένα και αδόμητα- σύνολα δεδομένων, ενώ η υιοθέτηση του RDF μοντέλου για την περιγραφή των δεδομένων μιας εφαρμογής παρέχει ευέλικτη μοντελοποίηση ενός δεδομένου τομέα ενδιαφέροντος και εύκολη ενημέρωση του σχήματος οργάνωσης των δεδομένων της εφαρμογής. Όσον αφορά στο χρήστη-καταναλωτή Συνδεδεμένων Δεδομένων, εκμεταλλευόμενος τη χρήση URI για τον μονοσήμαντο προσδιορισμό οντοτήτων και την ύπαρξη συνδέσμων μεταξύ σχετικών δεδομένων αποφεύγει σε μεγάλο βαθμό τις χρονοβόρες και επίπονες διαδικασίες εύρεσης αντιστοιχιών μεταξύ διαφορετικών συνόλων δεδομένων και ανακάλυψης νέων συνόλων δεδομένων.

Πλέον των προαναφερθέντων πλεονεκτημάτων τους, τα Συνδεδεμένα Δεδομένα κρίνονται επωφελή και για τις Βιβλιοθήκες, γεγονός που έχει διεθνώς αναγνωριστεί τα τελευταία χρόνια (Alemu, Stevens, Ross, & Chandler, 2012), (Byrne & Goddard, 2010), (Dunsire & Willer, 2010). Τα οφέλη που υπόσχονται τα Συνδεδεμένα Δεδομένα αφορούν όχι μόνο στις Βιβλιοθήκες ως οργανισμούς, αλλά και σε μεμονωμένα άτομα, όπως σε βιβλιοθηκονόμους, προγραμματιστές, ερευνητές και φοιτητές (Baker et al., 2011). Συγκεκριμένα, η αποτύπωση των μεταδεδομένων των Βιβλιοθηκών σε μορφή Συνδεδεμένων Δεδομένων καθιστά ευκολότερη την ανακάλυψη των τεκμηρίων της από ενδιαφερόμενους χρήστες. Όπως αναφέρθηκε και στην Ενότητα 1, ο Σημαιολογικός Ιστός αποτελεί μια επέκταση του Παγκόσμιου Ιστού, όπου τα δεδομένα αποκτούν δομημένη περιγραφή με αυστηρά καθορισμένο νόημα. Τα εν λόγω δομημένα δεδομένα, που συχνά προστίθενται σε υπάρχοντα έγγραφα Ιστού μέσω γλωσσών σηματολογικής επισημείωσης, όπως η RDFa και τα μικροδεδομένα, λαμβάνονται συνήθως υπόψη από τους αλγόριθμους υπολογισμού συνάφειας των μηχανών αναζήτησης και βελτιώνουν τη θέση στην οποία εμφανίζεται ένα έγγραφο στα αποτελέσματα μιας αναζήτησης. Αυτό έχει ως αποτέλεσμα την **αύξηση της ορατότητας των τεκμηρίων των Βιβλιοθηκών που δημοσιεύουν τους πόρους τους με τη μορφή Συνδεδεμένων Δεδομένων**, ενισχύοντας την παρουσία τους στον Παγκόσμιο Ιστό και προσελκύοντας νέους χρήστες. Επίσης, για τους ίδιους τους τελικούς χρήστες η ύπαρξη Συνδεδεμένων Δεδομένων προσφέρει προχωρημένες δυνατότητες πλοήγησης -μέσω σύνδεσης των τεκμηρίων των Βιβλιοθηκών με εξωτερικές υπηρεσίες όπως η Wikipedia - και ανακάλυψης των πόρων εκείνων που καλύπτουν τις πληροφοριακές ανάγκες τους.

Παράλληλα, το χαρακτηριστικό της εύκολης διασύνδεσης δεδομένων από διαφορετικές πηγές προσφέρει στις Βιβλιοθήκες τη **δυνατότητα να βελτιώσουν σημαντικά την ποιότητα των μεταδεδομένων τους**. Ένας δεδομένος πόρος μπορεί πλέον να περιγραφεί σε συνεργασία με άλλες Βιβλιοθήκες και να συνδεθεί με δεδομένα που προέρχονται από άλλες κοινότητες ή ακόμα και άτομα. Με αυτόν τον τρόπο, αναδεικνύεται για τις Βιβλιοθήκες η ευκαιρία να προσφέρουν στους χρήστες τους περιγραφές καλύτερης ποιότητας και λεπτομέρειας, μετασηματίζοντας τον παραδοσιακό τρόπο καταλογογράφησης, ο οποίος παράγει αυτόνομες μεμονωμένες περιγραφές τεκμηρίων, σε ένα συμμετοχικό μοντέλο επαναχρησιμοποίησης πληροφορίας που έχει δημοσιευθεί από εξωτερικούς οργανισμούς και άτομα. Αυτό το νέο μοντέλο καταλογογράφησης θα αποφέρει σημαντικά οφέλη και στο σύνολο των βιβλιοθηκονόμων, οι οποίοι θα είναι ελεύθεροι πλέον να εστιάζουν στον τομέα στον οποίο ειδικεύονται, χωρίς να χρειάζεται να τεκμηριώνουν εξ αρχής πόρους που έχουν ήδη τεκμηριωθεί επιτυχημένα από τρίτους.

Επίσης, τα Συνδεδεμένα Δεδομένα ανοίγουν νέους δρόμους για την **ανάπτυξη καινοτόμων εφαρμογών** που συνδυάζουν τα βιβλιοθηκονομικά δεδομένα με άλλα ανοικτά σύνολα δεδομένων και προσφέρουν μια διαφορετική οπτική στα πρώτα, τα οποία μέχρι σήμερα ήταν διαθέσιμα μόνο μέσω των ολοκληρωμένων βιβλιοθηκονομικών συστημάτων στα οποία είναι αποθηκευμένα. Ανοίγοντας τα δεδομένα τους και προσφέροντάς τα σε μια μηχαναγνώσιμη, εύκολα επεξεργάσιμη μορφή, οι Βιβλιοθήκες ενσωματώνουν τα δεδομένα τους στον Παγκόσμιο Ιστό, όπου μπορούν να επαναχρησιμοποιηθούν από εφαρμογές και να αξιοποιηθούν με μη αναμενόμενους τρόπους. Για παράδειγμα, ο συνδυασμός των δεδομένων δανεισμού μιας Βιβλιοθήκης με γεωχωρικά δεδομένα και δεδομένα δρομολογίων αστικών συγκοινωνιών θα μπορούσε να χρησιμοποιηθεί από μια εφαρμογή που προτείνει σε φοιτητές τη συντομότερη διαδρομή για την απόκτηση ενός συγκεκριμένου βιβλίου. Τέτοιες εφαρμογές αυξάνουν την ορατότητα των τεκμηρίων των Βιβλιοθηκών και συντελούν στην αύξηση της αναγνωρισιμότητάς τους.

Η εισαγωγή της τεχνολογίας των Συνδεδεμένων Δεδομένων στο χώρο των Βιβλιοθηκών μειώνει, επίσης, και την εξάρτηση από πρωτόκολλα και πρότυπα που συναντώνται αποκλειστικά στη βιβλιοθηκονομία (π.χ. Z39.50, SRU/SRW) και απαιτούν ειδικές γνώσεις από την πλευρά των προγραμματιστών. Αντίθετα, τα Συνδεδεμένα Δεδομένα βασίζονται σε ευρέως διαδεδομένες τεχνολογίες Ιστού, με αποτέλεσμα οι προγραμματιστές να έχουν ευκολότερο έργο, αλλά και οι ίδιες οι Βιβλιοθήκες να διαθέτουν περισσότερες επιλογές για τη στελέχωση του τεχνικού προσωπικού τους. Για τον ίδιο λόγο, **μειώνεται η εξάρτηση των Βιβλιοθηκών από εξειδικευμένα λογισμικά**, καθώς μπορούν πλέον να καταφεύγουν σε λογισμικά που χρησιμοποιούνται ευρύτατα στον Ιστό, πολλά εκ των οποίων είναι ελεύθερα και ανοικτού κώδικα.

#### 4. Παραγωγή Συνδεδεμένων Δεδομένων

Η παραγωγή και συντήρηση Συνδεδεμένων Δεδομένων είναι μια διαδικασία με ιδιαιτερότητες, απαιτεί δηλαδή την εφαρμογή μιας δομημένης μεθοδολογίας, όπως αυτήν του κύκλου ζωής Συνδεδεμένων Δεδομένων (Auer et al., 2012), όπου ορίζεται μια επαναληπτική διαδικασία οκτώ (8) σταδίων για την παραγωγή Συνδεδεμένων Δεδομένων υψηλής ποιότητας. Στην τρέχουσα ενότητα παρουσιάζεται η ακολουθία των σημαντικότερων βημάτων που απαιτούνται για τη μετατροπή βιβλιογραφικών -και όχι μόνο- δεδομένων σε Συνδεδεμένα Δεδομένα. Τα βήματα αυτά είναι:

1. Επιλογή και μοντελοποίηση των δεδομένων
2. Μετασχηματισμός σε μορφή RDF
3. Εύρεση συνδέσμων
4. Δημοσίευση Συνδεδεμένων Δεδομένων

**Βήμα 1<sup>ο</sup> (Επιλογή και μοντελοποίηση των δεδομένων):** Το στάδιο του σχεδιασμού, το οποίο περιλαμβάνει όλες τις απαραίτητες προεργασίες που προηγούνται της φάσης του μετασχηματισμού σε μορφή RDF, είναι ιδιαίτερης σημασίας για την παραγωγή Συνδεδεμένων Δεδομένων. Στο εν λόγω στάδιο πρέπει κατ' αρχήν να αναγνωριστούν τα σύνολα δεδομένων που θα καταστούν διαθέσιμα ως Συνδεδεμένα Δεδομένα. Για παράδειγμα, μια ακαδημαϊκή βιβλιοθήκη μπορεί να επιλέξει να μετατρέψει μόνο το βιβλιογραφικό της κατάλογο σε Συνδεδεμένα Δεδομένα και να συνεχίσει να διαθέτει τα μεταδεδομένα της γκρίζας βιβλιογραφίας αποκλειστικά μέσω του Αποθετηρίου της. Ένας κρίσιμος παράγοντας που επηρεάζει την απόφαση για το αν θα συμπεριληφθεί ένα σύνολο δεδομένων στη διαδικασία μετατροπής αφορά στα πνευματικά δικαιώματα των συγκεκριμένων δεδομένων, καθώς και στα θέματα προσωπικών δεδομένων που πιθανώς απορρέουν από αυτά. Αφενός είναι κρίσιμη η διασαφήνιση των πνευματικών δικαιωμάτων, έτσι ώστε να δηλωθεί ρητά η άδεια χρήσης η οποία θα διέπει τα RDF δεδομένα και να μην υπάρχουν ενδοιασμοί από τους δυνητικούς καταναλωτές τους, αφετέρου θα πρέπει να ληφθεί ιδιαίτερη μέριμνα ώστε να εξαιρεθούν από το στάδιο της μετατροπής ευαίσθητα δεδομένα.

Σε αυτό το στάδιο εντάσσεται και η μοντελοποίηση των επιλεγμένων δεδομένων σε RDF, η οποία περιλαμβάνει την επιλογή κατάλληλων οντολογιών που ταιριάζουν στη φύση των υπό εξέταση δεδομένων, καθώς και τον καθορισμό μιας ενιαίας μορφής που θα ακολουθούν τα URIs των νέων οντοτήτων. Στις σχετικές βιβλιογραφικές οντολογίες που ήδη αναφέρθηκαν στην Ενότητα 2 μπορούν, επίσης, να προστεθούν τα MADS/RDF, RDA, BIBFRAME, καθώς και η οικογένεια οντολογιών SPAR. Η επαναχρησιμοποίηση όσο το δυνατόν περισσότερων όρων από γνωστές θεματικές οντολογίες στον παραγόμενο RDF γράφο διευκολύνει πολύ την άμεση ολοκλήρωση με βιβλιογραφικά RDF δεδομένα από εξωτερικούς παρόχους και την κατανόηση της σημασιολογίας του RDF γράφου από χρήστες -ανθρώπους και μηχανές- που επιθυμούν να τον επεξεργαστούν. Σε ό, τι αφορά στην επιλογή ενιαίας πολιτικής για τη δομή των παραγόμενων URIs είναι επιθυμητό να ακολουθούνται τα επιτυχημένα παραδείγματα εφαρμογής και βέλτιστες πρακτικές από το χώρο του Σημασιολογικού Ιστού (Dodds & Davis, 2012; Sauermann & Cyganiak, 2008).

**Βήμα 2<sup>ο</sup> (Μετασχηματισμός σε μορφή RDF):** Μετά την επιλογή των συνόλων δεδομένων και τη θεωρητική μοντελοποίηση του RDF γράφου, ακολουθεί το πρακτικό στάδιο της μετατροπής των επιλεγμένων δεδομένων σε μορφή Συνδεδεμένων Δεδομένων. Το στάδιο αυτό είναι καθαρά προγραμματιστικό και η ροή εργασίας του διαφέρει ανάλογα με την εξεταζόμενη πηγή δεδομένων και το πρωτόκολλο επικοινωνίας με αυτή. Για παράδειγμα, στην περίπτωση που τα πρωτογενή δεδομένα είναι αποθηκευμένα σε μια σχεσιακή βάση δεδομένων, η εξαγωγή τους σε μορφή RDF μπορεί να γίνει με πολλά και διαφορετικά όργανα εργαλεία, όπως τα D2RQ (Bizer & Seaborne, 2004), Virtuoso RDF Views (Blakeley, 2007) και Ontop (Rodriguez-Muro, Kontchakov & Zakharyashev, 2013), ενώ στην περίπτωση που τα δεδομένα διατίθενται μέσω του πρωτοκόλλου OAI-PMH, μπορεί να χρησιμοποιηθεί το OAI2LOD εργαλείο (Haslhofer & Schandl, 2008). Αντίστοιχα, ανάλογα με τη μορφή των πρωτογενών δεδομένων, θα πρέπει να αναπτυχθεί η κατάλληλη προγραμματιστική διαδικασία για τη μετατροπή σε RDF γράφο.

Στις περισσότερες περιπτώσεις, η διαδικασία μετατροπής σε RDF θα πρέπει να εκτελείται ανά τακτά χρονικά διαστήματα, προκειμένου να εξασφαλιστεί ότι ο παραγόμενος RDF γράφος θα είναι όσο το δυνατόν επικαιροποιημένος, προκύπτοντας από τις πιθανές ενημερώσεις των πρωτογενών δεδομένων. Εξαιρεση αποτελεί η περίπτωση των σχεσιακών βάσεων δεδομένων, όπου είναι δυνατή μια προσέγγιση «νοητού» RDF γράφου, ο οποίος ανά πάσα στιγμή αντικατοπτρίζει τα τρέχοντα περιεχόμενα μιας βάσης δεδομένων (Spanos, Stavrou & Μίτρου, 2012).

**Βήμα 3<sup>ο</sup> (Εύρεση συνδέσμων):** Η μετατροπή σε RDF από μόνη της, χωρίς μέριμνα για την προσθήκη συνδέσμων προς άλλα σύνολα δεδομένων, οδηγεί απλώς στη δημιουργία απομονωμένων RDF γράφων. Το εν λόγω στάδιο χαρακτηρίζεται από την αυτόματη ή χειροκίνητη προσθήκη συνδέσμων, σύμφωνα με την 4<sup>η</sup> αρχή των Συνδεδεμένων Δεδομένων (Ενότητα 2), έτσι ώστε ο παραγόμενος RDF γράφος να συνδεθεί στο παγκόσμιο Σύννεφο Συνδεδεμένων Δεδομένων. Το πρώτο βήμα σε αυτό το στάδιο είναι η επιλογή των λεξιλογίων τιμών προς τα οποία θα αναζητηθούν σύνδεσμοι, όπως η DBPedia και τα υπόλοιπα λεξιλόγια που αναφέρθηκαν στην Ενότητα 2. Ο χειροκίνητος ορισμός συνδέσμων ενδείκνυται κυρίως για μικρά σύνολα δεδομένων με χαμηλό ρυθμό ενημέρωσης, ενώ για μεγαλύτερα σύνολα δεδομένων προτείνεται η αυτόματη ανακάλυψη αντιστοιχιών μέσω εργαλείων που προσπαθούν να εντοπίσουν λεξικολογικές και δομικές ομοιότητες μεταξύ οντοτήτων από διαφορετικά σύνολα δεδομένων. Δημοφιλή παραδείγματα των τελευταίων αποτελούν τα Silk (Volz, Bizer, Gaedke, & Kobilarov, 2009) και LIMES (Ngonga Ngomo & Auer, 2011).

**Βήμα 4<sup>ο</sup> (Δημοσίευση Συνδεδεμένων Δεδομένων):** Στόχος του τελευταίου σταδίου είναι η διαθεσιμότητα των παραχθέντων δεδομένων προκειμένου να μπορούν να επαναχρησιμοποιηθούν από χρήστες και εφαρμογές, σύμφωνα με την 3<sup>η</sup> αρχή των Συνδεδεμένων Δεδομένων που αναφέρθηκε στην Ενότητα 2. Βάσει αυτής, τα URIs των οντοτήτων που περιγράφονται στο νεο-παραχθέντα RDF γράφο θα πρέπει να είναι προσπελάσιμα μέσω του HTTP πρωτοκόλλου και να παρέχουν χρήσιμες πληροφορίες για την οντότητα στην οποία αντιστοιχούν, στη μορφή που επιθυμεί ο χρήστης. Βέλτιστες πρακτικές για τη δημοσίευση Συνδεδεμένων Δεδομένων έχουν καταγραφεί στη βιβλιογραφία (Heath & Bizer, 2011), ενώ έχουν αναπτυχθεί και εργαλεία, όπως το Pubby, τα οποία προσφέρουν μια οπτική Συνδεδεμένων Δεδομένων πάνω από οποιονδήποτε RDF γράφο. Επίσης, είναι επιθυμητό τα τελικά Συνδεδεμένα Δεδομένα να είναι διαθέσιμα σε διάφορες μορφές (π.χ. ύπαρξη τελικού σημείου εκτέλεσης ερωτημάτων και διάθεση συνολικού RDF γράφου για offline επεξεργασία), ώστε να είναι σε θέση να καλύψουν τις προτιμήσεις όσο το δυνατόν περισσότερων χρηστών.

## 5. Συμπεράσματα

Στο παρόν άρθρο επιχειρήθηκε μια σύντομη εισαγωγή και επισκόπηση της τεχνολογίας των Συνδεδεμένων Δεδομένων, η χρήση των εφαρμογών της οποίας -σε ποικίλους τομείς- αυξάνεται στις μέρες μας σε συχνότητα και ένταση. Επίσης, εξετάστηκαν συνοπτικά τα κύρια στάδια παραγωγής Συνδεδεμένων Δεδομένων υψηλής ποιότητας από πρωτογενή βιβλιογραφικά δεδομένα, ενώ αναφέρθηκαν και ορισμένα από τα ζητήματα που χρειάζεται να ληφθούν υπόψη κατά τη μετάβαση σ' αυτά. Ειδικά για τις Βιβλιοθήκες, τα Συνδεδεμένα Δεδομένα προσφέρουν μια άριστη ευκαιρία αναβάθμισης της ποιότητας των δεδομένων τους και απόκτησης από μέρους τους περισσότερων χρηστών στον Παγκόσμιο Ιστό, ενώ παράλληλα συνιστούν μια πολλά υποσχόμενη, επωφελή και για το προσωπικό των Βιβλιοθηκών, πρακτική. Ως εκ τούτου, και κινούμενες στην κατεύθυνση ανάλογων πρωτοβουλιών από διεθνείς Βιβλιοθήκες στην Ευρώπη και αλλού, οι ελληνικές Βιβλιοθήκες έχουν καθήκον να υιοθετήσουν τη φερέλπιδά νέα τεχνολογία των Συνδεδεμένων Δεδομένων.

## 6. Βιβλιογραφικές αναφορές

- Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2012). Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World*, 113(11/22), 549–570.
- Auer, S., Buhmann, L., Dirschl, C., Erling, O., Hausenblas, M., Isele, R., Lehmann, J., Martin, M., Mendes, P.N., van Nuffelen, B., Stadler, C., Tramp, S. & Williams, H. (2012). Managing the Life-Cycle of Linked Data with the LOD2 Stack. *Proceedings of the 11th International Semantic Web Conference (ISWC), 11-15 November 2012* (pp.1-16). Boston, USA.
- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., Panzer, M., Schneider, J., Singer, R., Summers, E., Waites, W., Young, J. & Zeng, M. (2011). *Library Linked Data Incubator Group Final Report*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 28-37.
- Berners-Lee, T. (2006). *Linked Data - Design Issues*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://www.w3.org/DesignIssues/LinkedData.html>

- Bizer, C. & Seaborne, A. (2004). D2RQ - Treating non-RDF Databases as Virtual RDF Graphs. Poster at the *3rd International Semantic Web Conference (ISWC)*, 7-11 November 2004. Hiroshima, Japan.
- Blakeley, C. (2007). *Virtuoso RDF Views – Getting Started Guide*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από [http://www.openlinksw.co.uk/virtuoso/Whitepapers/pdf/Virtuoso\\_SQL\\_to\\_RDF\\_Mapping.pdf](http://www.openlinksw.co.uk/virtuoso/Whitepapers/pdf/Virtuoso_SQL_to_RDF_Mapping.pdf)
- Brickley, D. & Guha, R.V. (2014). *RDF Schema 1.1*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://www.w3.org/TR/rdf-schema/>
- Byrne, G. & Goddard, L. (2010). The Strongest Link: Libraries and Linked Data. *D-Lib Magazine*, 16(11-12). doi: 10.1045/november2010-byrne
- Dodds, L. & Davis, I. (2012). *Linked Data Patterns*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://patterns.dataincubator.org/book/>
- Dunsire, G. & Willer, M. (2010). Initiatives to make standard library metadata models and structures available to the Semantic Web. *Proceedings of World Library and Information Congress: 76th IFLA General Conference and Assembly, 11-12 August 2010* (pp. 10-15). Gothenburg, Sweden.
- Haslhofer, B., & Schandl, B. (2008). *The OAI2LOD Server: Exposing OAI-PMH metadata as linked data*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://events.linkeddata.org/ldow2008/papers/03-haslhofer-schandl-oai2lod-server.pdf>
- Heath, T. & Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. San Rafael: Morgan & Claypool Publishers.
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P.F. & Rudolph, S. (2012). *OWL 2 Web Ontology Language Primer*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://www.w3.org/TR/owl2-primer>
- Malmsten, M. (2008). Making a library catalogue part of the Semantic Web. In J. Greenberg (Ed.), *Proceedings of the International Conference on Dublin Core and Metadata. Metadata for semantic and social applications, 22-26 September 2008* (pp. 146-152). Berlin: Universitätsverlag Göttingen.
- Ngonga Ngomo, A.C. & Auer, S. (2011). LIMES - a time-efficient approach for large-scale link discovery on the web of data. In T. Walsh (Eds.), *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, 16-11 July 2011 (pp. 2312-2317). AAAI Press. doi: 10.5591/978-1-57735-516-8/IJCAI11-385
- Rodríguez-Muro, M., Kontchakov, R., & Zakharyashev, M. (2013). Ontology-based data access: Ontop of databases. *Proceedings of the 12th International Semantic Web Conference (ISWC) Sydney, Australia, 21-25 October 2013* (pp. 558-573). Berlin Heidelberg: Springer. doi: 10.1007/978-3-642-41335-3\_35
- Sauermann, L. & Cyganiak, R. (2008). *Cool URIs for the Semantic Web*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://www.w3.org/TR/cooluris>
- Schreiber, G. & Raimond, Y. (2014). *RDF 1.1 Primer*. Ανακτήθηκε 14 Οκτωβρίου, 2014, από <http://www.w3.org/TR/rdf11-primer/>
- Spanos, D.-E., Stavrou, P. & Mitrou, N. (2012). Bringing relational databases into the Semantic Web: A survey, *Semantic Web Journal*, 3(2), 169–209. doi: 10.3233/SW-2011-0055
- Vila-Suero, D. & Gomez-Perez, A. (2013). datos.bne.es and MARiMbA: an insight into library linked data. *Library Hi Tech*, 31(4), 575-601. doi: 10.1108/LHT-03-2013-0031
- Volz, J., Bizer, C., Gaedke, M. & Kobilarov, G. (2009). Discovering and maintaining links on the web of data. *Proceedings of the 8th International Semantic Web Conference (ISWC)* (pp. 650-665). Washington DC, USA.





**Δρ. Δημήτριος-Εμμανουήλ Σπανός**

Ο Δρ. Δημήτριος-Εμμανουήλ Σπανός είναι ερευνητικός συνεργάτης στο Εθνικό Μετσόβιο Πολυτεχνείο. Απέκτησε το δίπλωμα του μηχανικού και το διδακτορικό δίπλωμα από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ το 2005 και 2013 αντίστοιχα, και μεταπτυχιακό δίπλωμα στη Διοίκηση Επιχειρήσεων (MBA) από το Οικονομικό Πανεπιστήμιο Αθηνών το 2007. Είναι μέλος της Κεντρικής Ομάδας Υποστήριξης του Συνδέσμου Ελληνικών Ακαδημαϊκών Βιβλιοθηκών και μέλος της ερευνητικής ομάδας Επικοινωνιών Πολυμέσων και Τεχνολογιών Παγκόσμιου Ιστού του ΕΜΠ. Τα ερευνητικά του ενδιαφέροντα εντοπίζονται στα πεδία του Σημασιολογικού Ιστού, Συνδεδεμένων Δεδομένων και αναπαράστασης γνώσης με έμφαση στην αντιστοιχία σχεσιακών βάσεων δεδομένων με οντολογίες και τη σημασιολογική αναζήτηση σε ψηφιακές βιβλιοθήκες. Εργασίες του έχουν παρουσιαστεί σε διεθνή συνέδρια και δημοσιευτεί σε διεθνή περιοδικά με κριτές, έχοντας συγκεντρώσει πάνω από 100 ετεροαναφορές. Είναι μέλος του Τεχνικού Επιμελητηρίου Ελλάδας.