

# DIAS: Ένα σύστημα Ιδρυματικού Αποθετηρίου βασισμένο στο πρότυπο FRBR

Νεκτάριος Γιολδάσης, Φώτης Καζάσης, Γιώργος Ανέστης, Νίκος Παππάς

{nektarios,fotis,ganest,nikos@ced.tuc.gr}

Πολυτεχνείο Κρήτης

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων και Εφαρμογών Πολυμέσων

## Περίληψη

Το σύστημα DIAS (Digital Information Archival System) είναι μια υποδομή Ιδρυματικού Αποθετηρίου που σχεδιάστηκε και αναπτύχθηκε από το Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων (MUSIC) του Πολυτεχνείου Κρήτης. Αξιοποιεί έννοιες από το πρότυπο FRBR για να μοντελοποιήσει και να περιγράψει πνευματικά έργα και τις διάφορες εκφράσεις τους, επιτρέποντας ταυτόχρονα τον καθορισμό του σχήματος περιγραφικών μετά-δεδομένων αλλά και συγκεκριμένων εξειδικεύσεων του ανάλογα με τη συλλογή ή τον τύπο των τεκμηρίων. Το σύστημα DIAS μπορεί να παραμετροποιηθεί έτσι ώστε να υποστηρίζει εξειδικευμένες ροές εργασίας σε οποιαδήποτε συμβάν κατά τη διάρκεια ζωής ενός τεκμηρίου. Είναι πλήρως διαλειτουργικό με τρίτα συστήματα, υποστηρίζοντας τα πρωτόκολλα OAI-PMH, SWORD και Open Search, ενώ μπορεί να εξαγάγει τα μετά-δεδομένα σε μια (επεκτάσιμη) σειρά από μορφότυπα. Επίσης, το σύστημα DIAS υποστηρίζει την παροχή προσωποποιημένης διάχυσης περιεχομένου, δίνοντας στους χρήστες τη δυνατότητα να δηλώνουν τις προτιμήσεις τους και να λαμβάνουν ειδοποιήσεις κάθε φορά που δημοσιεύεται κάποιο τεκμήριο το οποίο ικανοποιεί τις προτιμήσεις αυτές.

Στο άρθρο αυτό παρουσιάζουμε το σύστημα DIAS και τα βασικά του χαρακτηριστικά, δίνοντας έμφαση στο πως το πρότυπο FRBR μπορεί να χρησιμοποιηθεί για να περιγράψει γκριζα βιβλιογραφία και γενικά επιστημονικές εργασίες. Συζητάμε επίσης, πως το σύστημα DIAS αξιοποιεί τις αρχές μοντελοποίησης του προτύπου SKOS για να διασυνδέσει τα τεκμήρια με δημοσιευμένους πόρους σε θησαυρούς, οντολογίες, συστήματα ταξινόμησης, κλπ. ή για να υποστηρίξει τη δυνατότητα παροχής πολλαπλών τιμών σε ένα γνώρισμα τεκμηρίου.

## Εισαγωγή

Το σύστημα DIAS αναπτύχθηκε στο πλαίσιο του έργου "Υπηρεσίες Ψηφιακής Βιβλιοθήκης Ανοικτής Πρόσβασης Πολυτεχνείου Κρήτης" του Πολυτεχνείου Κρήτης για την υλοποίηση του υπο-έργου "Υπηρεσία Δημιουργίας Ιδρυματικού Αποθετηρίου/Ψηφιακής Βιβλιοθήκης Ανοικτής Πρόσβασης" και χρηματοδοτήθηκε από το πρόγραμμα "Ψηφιακή Σύγκλιση". Παρ' ότι ο βασικός σκοπός του έργου ήταν η υλοποίηση του Ιδρυματικού Αποθετηρίου του Πολυτεχνείου Κρήτης, το σύστημα DIAS σχεδιάστηκε και υλοποιήθηκε έτσι ώστε να μπορεί να παραμετροποιηθεί κατάλληλα και να αξιοποιηθεί σε πολλές περιπτώσεις όπου υπάρχουν ανάγκες αρχειοθέτησης, διατήρησης και διάχυσης ψηφιακού υλικού. Είναι μια πρότυπη υλοποίηση του μοντέλου αναφοράς για ανοικτά πληροφοριακά συστήματα αρχειοθέτησης (Open Archival Information Systems -Reference Model [1]) του διεθνούς οργανισμού OCLC το οποίο λειτουργεί ως σημείο αναφοράς σχεδιασμού και υλοποίησης συστημάτων αρχειοθέτησης. Το μοντέλο αυτό καθορίζει διακριτές λειτουργικές μονάδες για όλα τα στάδια επεξεργασίας του υλικού σε μια τυπική ροή εργασιών αρχειοθέτησης (archival workflow) από την τροφοδοσία (ingest), την αρχειοθέτηση (archival), τη διαχείριση (management), μέχρι και την προσπέλαση/διάχυση (access/dissemination). Καθορίζει επίσης τα διάφορα είδη μεταδεδομένων (περιγραφικά,

δομικά, διαχειριστικά, προέλευσης, πρόσβασης, κλπ.) τα οποία πρέπει να διατίθενται ή να καταγράφονται σε κάθε στάδιο. Είναι ωστόσο σημαντικό να τονίσουμε ότι το μοντέλο καθορίζει μόνο τα είδη και όχι το ακριβές σχήμα των μεταδεδομένων, το οποίο αφήνεται στην ευχέρεια των σχεδιαστών ενός συγκεκριμένου συστήματος αρχαιοθήκης.

## Μοντελοποίηση μεταδεδομένων στο σύστημα DIAS

Από όλα τα είδη των μεταδεδομένων που καθορίζει το OAIIS RM, καθοριστικής σημασίας είναι τα περιγραφικά μεταδεδομένα. Αυτό γιατί τόσο κατά την τροφοδοσία του συστήματος, όσο και κατά την προσπέλαση υλικού από αυτό, οι τελικοί χρήστες (παραγωγοί και καταναλωτές αντίστοιχα σύμφωνα με το OAIIS RM) ανταλλάσσουν πακέτα πληροφορίας (information packages) τα οποία αποτελούνται κατά κύριο λόγο από περιγραφικά μεταδεδομένα τεκμηρίων.

Τα περισσότερα παραδοσιακά πρότυπα μεταδεδομένων (MARC21[2], UNIMARC[3], MODS[4], Dublin Core[5], LMAP[6], LOM[7], κ.α.) ακολουθούν τη λογική των εγγραφών περιγραφής (record-based formats), όπου για κάθε αντικείμενο υπάρχει μια μοναδική εγγραφή με πληροφορία γιαυτό.

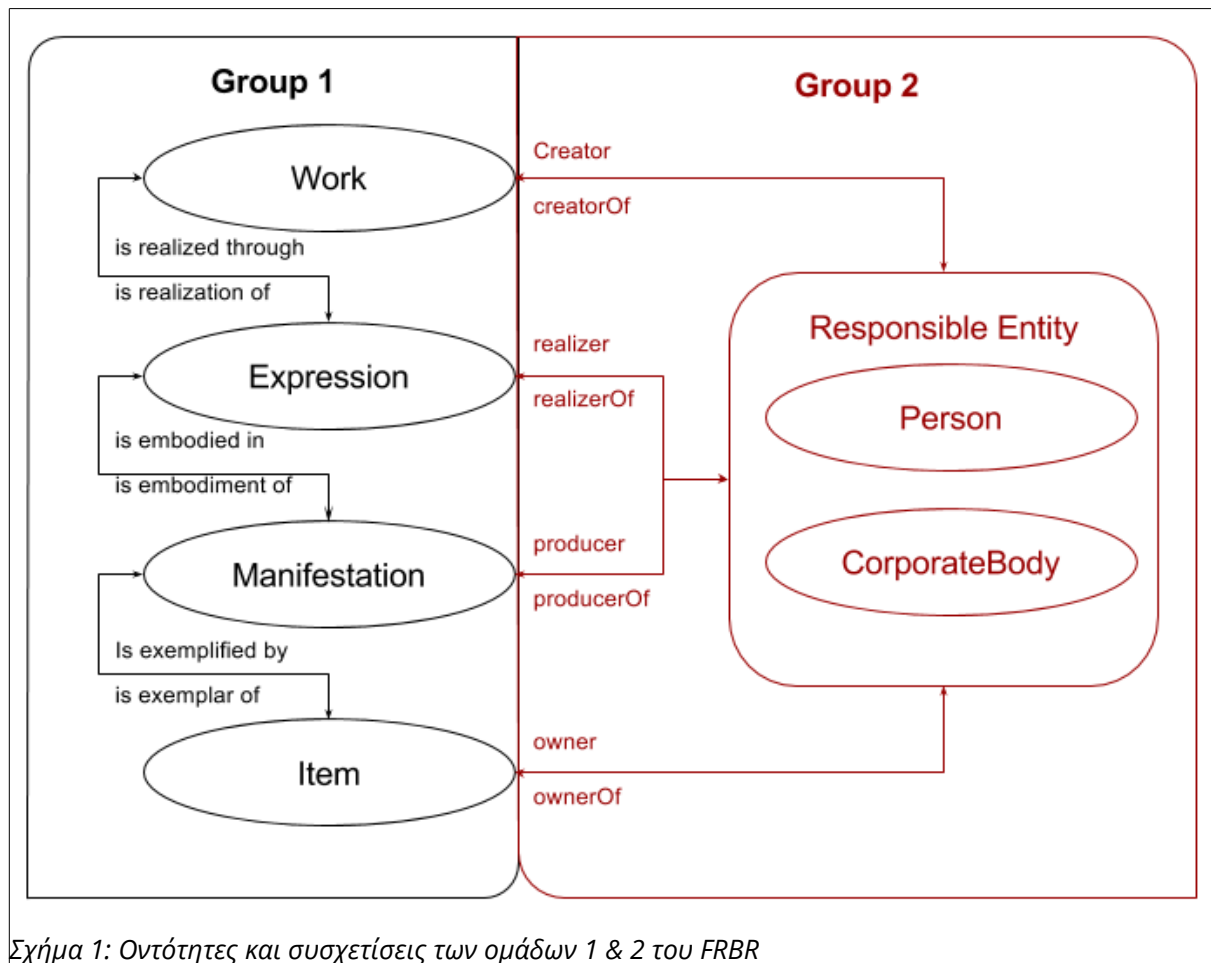
Η μοντελοποίηση περιγραφικών μεταδεδομένων στο σύστημα DIAS βασίζεται στο πρότυπο FRBR (Functional Requirements for Bibliographic Records)[8]. Το σύστημα παρέχει μηχανισμό δημιουργίας μοντέλων (δηλαδή ένα μετά-μοντέλο) περιγραφικών μεταδεδομένων ο οποίος επιτρέπει την περιγραφή αντικειμένων σε διαφορετικά επίπεδα αφαίρεσης αλλά και χρήση εξειδικευμένων γνωρισμάτων περιγραφής (σε κάθε ένα από αυτά τα επίπεδα) ανάλογα με τις εκάστοτε ανάγκες.

Το πρότυπο FRBR προτείνει ένα μοντέλο οντοτήτων-συσχετίσεων (entity relationship model) προκειμένου να παράσχει μεγάλη ευελιξία στις δυνατότητες περιγραφής ενός έργου (π.χ. περιγραφή ενός έργου σε διαφορετικά επίπεδα αφαίρεσης) αλλά και να επιτρέψει την ύπαρξη αυθύπαρκτων οντοτήτων οι οποίες συσχετίζονται κατά το δοκούν για την περιγραφή έργων στα οποία εμπλέκονται. Οι βασικές έννοιες του προτύπου ομαδοποιούνται σε 3 βασικές ομάδες. Η πρώτη ομάδα περιλαμβάνει τις οντότητες και συσχετίσεις που μοντελοποιούν ένα έργο σε διαφορετικά επίπεδα αφαίρεσης. Η δεύτερη ομάδα περιλαμβάνει οντότητες και συσχετίσεις οι οποίες μοντελοποιούν την έννοια των εμπλεκόμενων (οργανισμών ή ανθρώπων) οντοτήτων που έχουν κάποια σχέση υπευθυνότητας (δημιουργοί, εκδότες, κάτοχοι, κλπ) με το έργο (σε κάποιο επίπεδο αφαίρεσης). Η τρίτη ομάδα τέλος, περιλαμβάνει τις οντότητες και συσχετίσεις οι οποίες μοντελοποιούν τη θεματική κατηγοριοποίηση των έργων.

Το Σχήμα 1 δείχνει τις δύο πρώτες ομάδες του προτύπου FRBR. Οι οντότητες και οι συσχετίσεις της Ομάδας 1 ορίζουν μια ιεραρχική δομή για την περιγραφή ενός έργου σε διάφορα επίπεδα αφαίρεσης. Συγκεκριμένα:

- **Έργο (Work):** Είναι μια αφηρημένη οντότητα που αναπαριστά ένα πνευματικό ή καλλιτεχνικό δημιούργημα (π.χ. "Ο Άρχοντας των Δακτυλιδιών") σε υψηλό επίπεδο αφαίρεσης τέτοιο ώστε το δημιούργημα αυτό να μην σχετίζεται με κανένα φυσικό αντικείμενο.
- **Έκφραση (Expression):** Είναι μια πνευματική/καλλιτεχνική έκφραση του έργου σε μια συγκεκριμένη μορφή (π.χ. Το βιβλίο "Ο Άρχοντας των Δακτυλιδιών" σε κάποια γλώσσα, ή η ταινία "Ο Άρχοντας των Δακτυλιδιών").
- **Υλοποίηση (Manifestation):** Είναι μια συγκεκριμένη υλοποίηση (παραγωγή) μιας έκφρασης του έργου (π.χ. το βιβλίο "Ο Άρχοντας των Δακτυλιδιών" στα αγγλικά και σε έκδοση με σκληρό εξώφυλλο)
- **Αντίτυπο (Item):** Είναι ένα συγκεκριμένο αντίτυπο μιας υλοποίησης το οποίο βρίσκεται σε κάποια συγκεκριμένη θέση ή κατέχεται από κάποιο συγκεκριμένο άτομο, κλπ.

Οι οντότητες Έργο και Έκφραση, περιγράφουν την πνευματική διάσταση ενός έργου, ενώ οι οντότητες Υλοποίηση και Αντίτυπο περιγράφουν την φυσική/τεχνική διάστασή του.

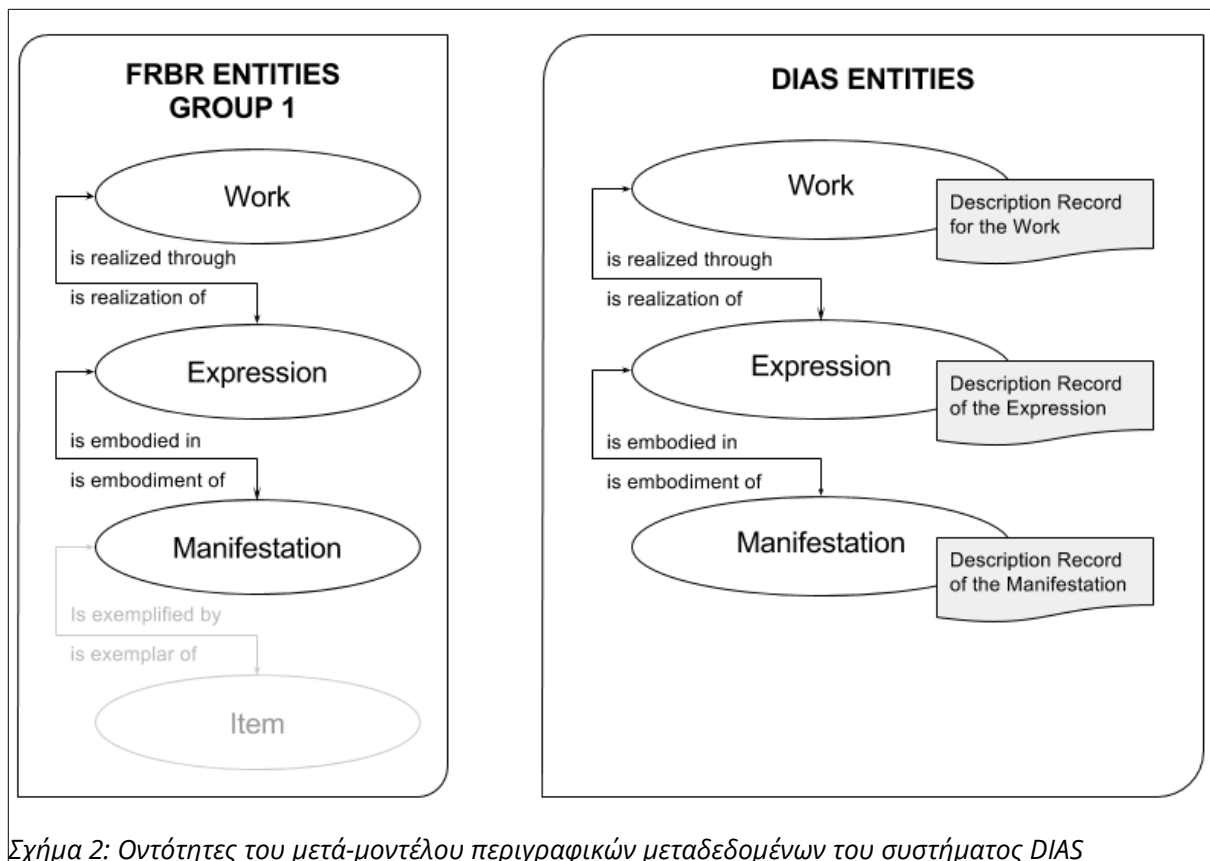


Σχήμα 1: Οντότητες και συσχετίσεις των ομάδων 1 & 2 του FRBR

## Οντότητες και δομικές συσχετίσεις

Το μέτα-μοντέλο περιγραφικών μεταδεδομένων του συστήματος DIAS συνδυάζει τη σχεσιακή θεώρηση που προτείνεται από το FRBR με τη λογική των εγγραφών περιγραφής για κάθε οντότητα. Έτσι υιοθετεί τις οντότητες και συσχετίσεις της ομάδας 1 του προτύπου FRBR επιτρέποντας μια ιεραρχική/αφαιρετική περιγραφή των πνευματικών δημιουργημάτων και ειδικότερα των ακαδημαϊκών εργασιών (scholarly works) όπως προτείνεται και από το μοντέλο SWAP (Scholarly Works Application Profile)[9]. Ωστόσο, το σύστημα DIAS αφορά σε ηλεκτρονικές εκδόσεις πνευματικών έργων (eprints) και δεν είναι στους σκοπούς του η καταγραφή και περιγραφή των (ψηφιακών) αντιτύπων των έργων. Έτσι, δεν χρησιμοποιεί την οντότητα *Αντίτυπο* και φυσικά τις συσχετίσεις οι οποίες την αφορούν.

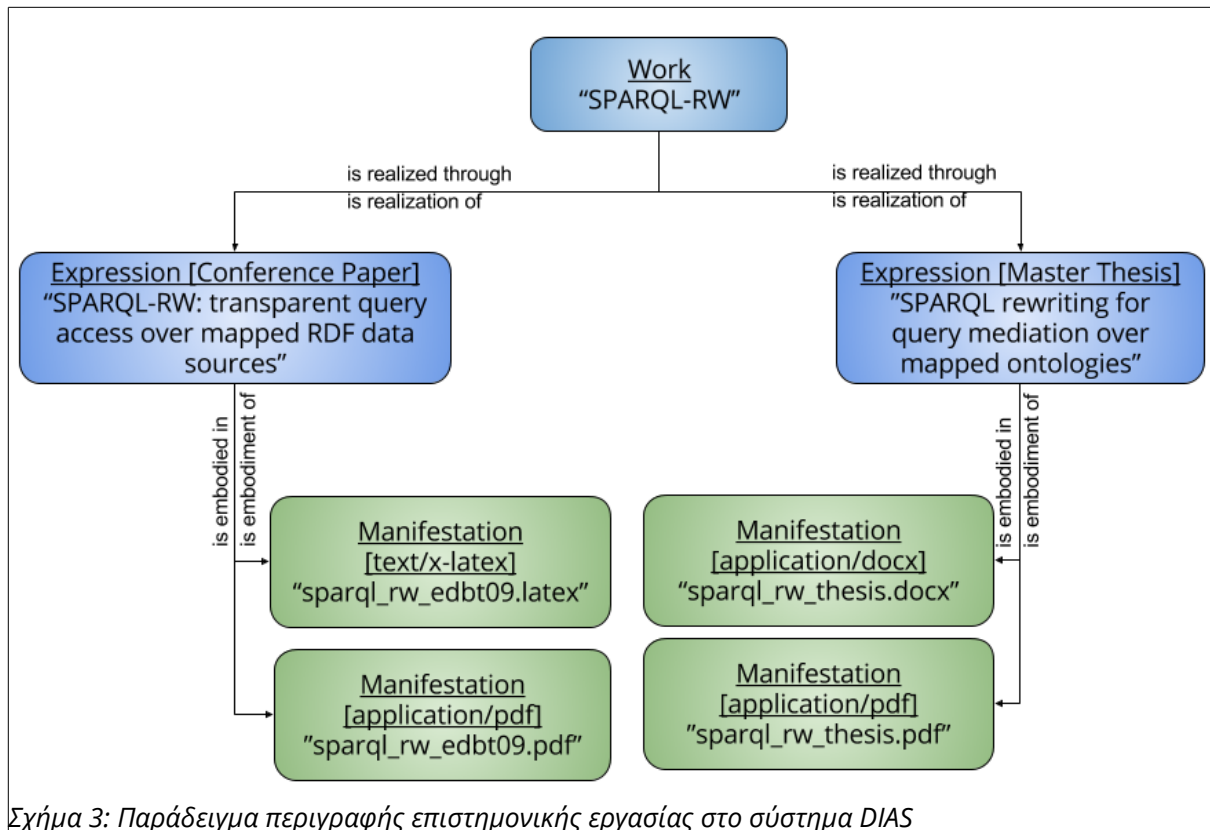
Το Σχήμα 2 δείχνει τις οντότητες και τις δομικές συσχετίσεις τους στο σύστημα DIAS και την αντιστοίχσή τους με αυτή του προτύπου FRBR.



Σχήμα 2: Οντότητες του μετά-μοντέλου περιγραφικών μεταδεδομένων του συστήματος DIAS

Ένα **Έργο** (Work) είναι μια αφηρημένη οντότητα η οποία αναπαριστά ένα ακαδημαϊκό (ερευνητικό/επιστημονικό/εκπαιδευτικό) έργο ως πνευματικό πόνημα. Ένα τέτοιο έργο μπορεί να λαμβάνει διάφορες μορφές (Expressions) κάθε μια από τις οποίες είναι μια τεκμηρίωση του Έργου (ένα **Τεκμήριο**). Κάθε τεκμήριο μπορεί να εμπεριέχεται με τη σειρά του σε ένα -ηλεκτρονικό- **Αρχείο** (Manifestation), ή και περισσότερα.

Το (υπαρκτό) παράδειγμα που παρουσιάζεται στο Σχήμα 3 παρουσιάζει την εκφραστικότητα της ακολουθούμενης θεώρησης. Το έργο "SPARQL Rewriting" είναι μια επιστημονική εργασία (ένα πνευματικό έργο) η οποία πραγματεύεται και προτείνει μια μεθοδολογία για την αυτόματη "μετάφραση" ερωτημάτων σε περιβάλλον σημασιολογικού ιστού. Η (ερευνητική) εργασία αυτή έχει λάβει τη μορφή (τύπο) μιας Δημοσίευσης σε Συνέδριο με τον τίτλο "SPARQL-RW: transparent query access over mapped RDF data sources" αλλά και τη μορφή (τύπο) μιας Μεταπτυχιακής Διατριβής στο Πολυτεχνείο Κρήτης με τίτλο "SPARQL rewriting for query mediation over mapped ontologies". Και τα δύο αυτά **Τεκμήρια** αποτελούν τεκμηρίωση (έκφραση) του ίδιου πνευματικού έργου με λιγότερες ή περισσότερες λεπτομέρειες. Κάθε ένα από τα τεκμήρια αυτά εμπεριέχεται σε δύο ηλεκτρονικά **Αρχεία**: ένα αρχείο σε επεξεργάσιμη μορφή και ένα αρχείο σε μη επεξεργάσιμη μορφή.



Σχήμα 3: Παράδειγμα περιγραφής επιστημονικής εργασίας στο σύστημα DIAS

## Γνωρίσματα και σημασιολογική συσχέτιση οντοτήτων

Για κάθε μία από τις οντότητες του μετά-μοντέλου, το σύστημα DIAS διατηρεί μια πλήρη εγγραφή με τα γνωρίσματά της. Διαφορετικές εγκαταστάσεις του συστήματος DIAS μπορούν να ορίσουν διαφορετικά μοντέλα μεταδεδομένων (συγκεκριμενοποιώντας -instantiating- το μετά-μοντέλο) δημιουργώντας γνωρίσματα εγγραφής που θεωρούνται σημαντικά για κάθε οντότητα στο πλαίσιο των απαιτήσεων του εκάστοτε αποθετηρίου. Για παράδειγμα κατά την ανάπτυξη του Ιδρυματικού Αποθετηρίου του Πολυτεχνείου Κρήτης δημιουργήθηκε ένα συγκεκριμένο μοντέλο μεταδεδομένων ορίζοντας για κάθε οντότητα γνωρίσματα τέτοια ώστε να καλύπτονται οι ανάγκες του ιδρύματος, όσο και οι προδιαγραφές του Συνδέσμου Ελληνικών Ακαδημαϊκών Βιβλιοθηκών σχετικά με ένα ελάχιστο κοινό σχήμα μεταδεδομένων για όλα τα ιδρυματικά αποθετήρια της Ελλάδας.

Κατά τη δημιουργία ενός μοντέλου μεταδεδομένων, ο δημιουργός του μπορεί να επιλέξει για κάθε γνώρισμα το όνομά του, την πληθικότητά του, και τον τύπο του. Ο καθορισμός του τύπου ενός γνωρίσματος είναι πολύ σημαντικός καθώς επηρεάζει κατά πολύ την ποιότητα των μεταδεδομένων μιας οντότητας.

Το σύστημα DIAS υποστηρίζει δύο είδη γνωρισμάτων εγγραφής: Τα **γνωρίσματα πρωταρχικού τύπου** (primitive elements), και τα **γνωρίσματα αναφοράς** (reference elements). Στα γνωρίσματα πρωταρχικού τύπου περιλαμβάνονται συμβολοσειρές, αριθμοί, ημερομηνίες, τιμές αληθείας, καθώς και αλφαριθμητικά με γλωσσική επισήμανση (εκτός δηλαδή από την τιμή του γνωρίσματος, καταγράφεται και ο κωδικός της γλώσσας στην οποία αυτή παρέχεται).

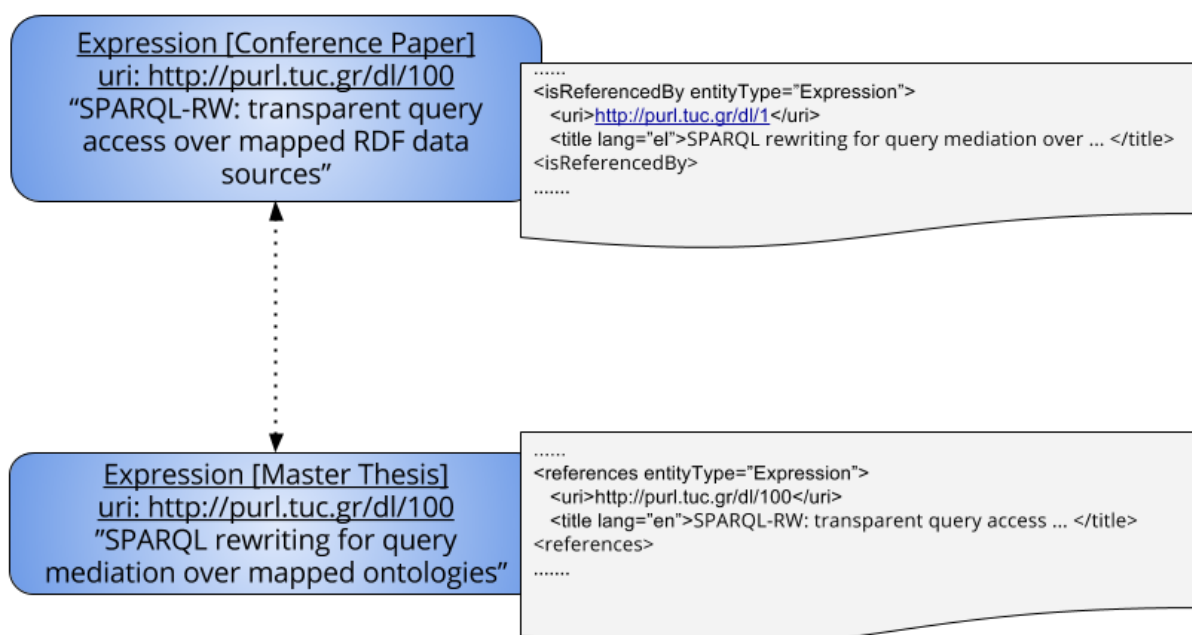
Τα γνωρίσματα αναφοράς, συσχετίζουν την περιγραφόμενη οντότητα με μια άλλη οντότητα είτε εντός του συστήματος, είτε έξω από το σύστημα. Για την μοντελοποίηση γνωρισμάτων αναφοράς, το σύστημα DIAS χρησιμοποιεί δομή εμπνευσμένη από την δομή της "έννοιας"

(concept) στο πρότυπο SKOS (Simple Knowledge Organization System)[10] όπου κάθε έννοια περιγράφεται ως μια συλλογή από (εναλλακτικές ή μεταφρασμένες) ονομασίες. Αυτή η δομή υιοθετείται για γνωρίσματα αναφοράς και επεκτείνεται με άλλα χαρακτηριστικά όπως θα δείξουμε παρακάτω.

### Συσχετίσεις μεταξύ οντοτήτων συστήματος

Εκτός από τις δομικές συσχετίσεις μεταξύ των οντοτήτων του μοντέλου, υπάρχουν και άλλες οριζόντιες συσχετίσεις σε κάθε επίπεδο. Ένα Έργο μπορεί να “είναι μέρος” από, να “επεκτείνει”, να “έχει ως θέμα”, κλπ ένα άλλο έργο. Αντίστοιχα ένα Τεκμήριο μπορεί να “αναφέρει” ένα άλλο τεκμήριο (όχι απαραίτητα στο πλαίσιο του ίδιου έργου), ή να “αποτελεί μετάφραση” ενός τεκμηρίου (του ίδιου έργου), κλπ. Οριζόντιες συσχετίσεις υπάρχουν τέλος και στο επίπεδο των Αρχείων αλλά μόνο στο πλαίσιο του ίδιου Τεκμηρίου. Έτσι, ένα αρχείο μπορεί να “αποτελεί μετατροπή”, να “προηγείται” ή να “έπεται” ενός άλλου αρχείου κλπ. Για παράδειγμα, κάθε κεφάλαιο ενός τεκμηρίου της μορφής “Διδακτορική Διατριβή” μπορεί να εμπεριέχεται σε ξεχωριστό αρχείο. Σε μια τέτοια περίπτωση, είναι σημαντικό να οριστεί μια σειρά προσπέλασης αυτών των αρχείων (μέσω συσχετίσεων “προηγείται σε σειρά” και “έπεται σε σειρά” του αρχείου).

Για τη σημασιολογική συσχέτιση των οντοτήτων μεταξύ τους, το μέτα-μοντέλο προβλέπει τον τύπο γνωρίσματος “**Σχετιζόμενη Οντότητα**”. Ένα γνώρισμα αυτού του τύπου, είναι ένα γνώρισμα αναφοράς με σύνθετη δομή η οποία περιλαμβάνει: το μοναδικό αναγνωριστικό (URI) της σχετιζόμενης οντότητας, το είδος της (Έργο/Τεκμήριο/Αρχείο) και τον τίτλο της σε κάποια γλώσσα. Οι συσχετίσεις μεταξύ των οντοτήτων του συστήματος DIAS είναι αμφίδρομες. Το μοντέλο μεταδεδομένων μιας οντότητας πρέπει δηλαδή να ορίζει ζευγάρια γνωρισμάτων αυτού του τύπου. Για παράδειγμα αν σε ένα σχήμα μεταδεδομένων θέλουμε να ορίσουμε ότι τα τεκμήρια μπορούν να αναφέρουν άλλα τεκμήρια, τότε θα πρέπει να ορίσουμε ένα ζεύγος γνωρισμάτων (τύπου “σχετιζόμενη οντότητα”) για την οντότητα Τεκμήριο: το γνώρισμα “αναφέρει” και το γνώρισμα “αναφέρεται από”. Έτσι αν στο παράδειγμα του Σχήματος 3 υποθέσουμε ότι το τεκμήριο της Μεταπτυχιακής Διατριβής αναφέρει το τεκμήριο της Δημοσίευσης σε Συνέδριο, οι εγγραφές τους θα περιλαμβάνουν τιμές για τα γνωρίσματα “αναφέρει” και “αναφέρεται από” όπως φαίνεται στο Σχήμα 4.



Σχήμα 4: Σημασιολογική συσχέτιση οντοτήτων του συστήματος

## Συσχετίσεις με οντότητες εκτός συστήματος

Το σύστημα DIAS δεν έχει υιοθετήσει τις οντότητες των Ομάδων 2 και 3 του προτύπου FRBR. Αντ' αυτού, και ακολουθώντας την πρακτική των διασυνδεδεμένων δεδομένων (linked data), χρησιμοποιεί κατά κόρον γνωρίσματα αναφοράς σε δημοσιευμένους πόρους (published resources). Τέτοιοι πόροι μπορεί να είναι έννοιες από online θησαυρούς (thesauri) ή συστήματα ταξινόμησης (classification schemes), μπορεί να είναι οντότητες από αρχεία οντοτήτων (authority files), ή γενικά πόροι (RDF resources) από το περιβάλλον των ανοικτών διασυνδεδεμένων δεδομένων.

Για την καταγραφή μιας αναφοράς προς έναν δημοσιευμένο πόρο, το μοντέλο μεταδεδομένων του συστήματος DIAS ορίζει τον τύπο γνωρίσματος **“Αναφορά σε Λεξικό”** (Vocabulary Reference). Ένα γνώρισμα αυτού του τύπου είναι γνώρισμα αναφοράς, με δομή η οποία περιλαμβάνει:

- το παγκόσμια μοναδικό αναγνωριστικό του πόρου με τη μορφή URI (unique resource identifier)
- το λογικό όνομα του λεξικού ή του συστήματος εννοιών, ή του αρχείου οντοτήτων, ή του θησαυρού, κλπ. σε μορφή αλφαριθμητικού
- τον κωδικό (σε μορφή αλφαριθμητικού) του πόρου ο οποίος είναι μοναδικός εντός των ορίων του συστήματος ορισμού (concept scheme, vocabulary, authority file, κλπ)
- την προτιμώμενη ονομασία του πόρου σε φυσική γλώσσα.
- μια σειρά από ονομασίες (labels) του πόρου σε φυσική γλώσσα με γλωσσική επισήμανση.

Μία επέκταση του τύπου γνωρίσματος “Αναφορά σε Λεξικό” είναι ο τύπος **“Συντελεστής”** (contributor) ο οποίος επεκτείνει την “Αναφορά σε Λεξικό” με πληροφορίες για το ρόλο που είχε η αναφερόμενη οντότητα για την αναφέρουσα. Οι ρόλοι είναι και αυτοί έννοιες/οντότητες οι οποίες είναι δημοσιευμένες σε κάποιο λεξικό, ή σύστημα εννοιών. Έτσι στην πράξη, γνωρίσματα αυτού του τύπου καταγράφουν μια τριαδική σχέση μεταξύ μιας οντότητας του μοντέλου, μιας οντότητας του εξωτερικού κόσμου, και της έννοιας/οντότητας του ρόλου.

Τέλος, πέραν των συσχετίσεων με δημοσιευμένες οντότητες η δομή των γνωρισμάτων αναφοράς χρησιμοποιείται και για γνωρίσματα των οποίων η τιμή είναι κάποια μη δημοσιευμένη οντότητα. Με τη χρήση γνωρισμάτων αυτού του τύπου, μπορεί για παράδειγμα να καταγραφεί η ονομασία ενός συντελεστή τεκμηρίου τόσο στα Ελληνικά όσο και στα Αγγλικά διατηρώντας ταυτόχρονα την πληροφορία ότι πρόκειται για το ίδιο πρόσωπο.

Η μοντελοποίηση των γνωρισμάτων αναφοράς με τη δομή των εννοιών SKOS καθιστά την εγγραφή περιγραφικών μεταδεδομένων μιας οντότητας αυθύπαρκτη (self-contained). Δεν χρειάζεται δηλαδή να ανατρέξει κανείς στην αναφερόμενη οντότητα για να βρει προφανείς πληροφορίες γι' αυτή (π.χ. ονομασίες). Επίσης, δίνει το πλεονέκτημα, ότι πληροφορίες της αναφερόμενης οντότητας χρησιμοποιούνται στη δεικτοδότηση της αναφέρουσας. Έτσι, για παράδειγμα, ένα ερώτημα που περιλαμβάνει τον όρο “query mediation” στο παράδειγμα του Σχήματος 4 θα επιστρέψει τόσο τη διπλωματική εργασία όσο και τη δημοσίευση σε συνέδριο.

## Κύρια λειτουργικά χαρακτηριστικά του συστήματος DIAS

Το σύστημα DIAS διαθέτει όλα τα λειτουργικά χαρακτηριστικά ενός μοντέρνου συστήματος αποθετηρίου. Μερικά από τα πιο καινοτόμα λειτουργικά χαρακτηριστικά του συστήματος, ανά κατηγορία, είναι:

- **Οργάνωση Περιεχομένου.** Το σύστημα DIAS υποστηρίζει μια ιεραρχική οργάνωση κοινοτήτων προκειμένου να μοντελοποιήσει τη δομή σύνθετων οργανισμών όπως είναι οι Ακαδημαϊκές Μονάδες. Υποστηρίζει τη δημιουργία πρότυπων συλλογών (abstract collections) (π.χ. "Διπλωματικές Εργασίες") οι οποίες όταν συσχετίζονται με μια κοινότητα (π.χ. "Σχολή HMMY") δημιουργούν μια συγκεκριμένη (concrete) συλλογή (π.χ. "Διπλωματικές εργασίες σχολής HMMY"). Μια πρότυπη συλλογή μπορεί να ορίζει τους τύπους τεκμηρίων που μπορεί να περιέχει. Τα τεκμήρια μιας συγκεκριμένης συλλογής, εμφανίζονται στις αντίστοιχες συλλογές όλων των υπερ-κοινοτήτων.
- **Χρήστες & Ρόλοι.** Υποστηρίζεται καθορισμός και απόδοση ρόλων σε μεμονωμένους χρήστες του συστήματος ή ομάδες χρηστών. Υποστηρίζεται επίσης η δυνατότητα αυτόματης δημιουργίας ομάδων χρηστών για κάθε κοινότητα και συλλογή, καθώς και η αυτόματη απόδοση ρόλων σε χρήστες σε συνεργασία με καταλόγους LDAP[11].
- **Ροές Εργασιών:** Το σύστημα DIAS υποστηρίζει τόσο την ελεγχόμενη, όσο και την ελεύθερη κατάθεση (χωρίς έγκριση) υλοποιώντας τις σχετικές ροές εργασιών. Πέραν αυτού, διαθέτει έναν μηχανισμό επέκτασης αυτών των ροών σύμφωνα με τις ανάγκες που καθορίζονται από την πολιτική διατήρησης ή λειτουργίας ενός αποθετηρίου. Συγκεκριμένα, επιτρέπει τον καθορισμό εργασιών που θα εκτελεστούν (σύγχρονα ή ασύγχρονα) κατά την ολοκλήρωση των δραστηριοτήτων μιας ροής εργασιών. Ο Καθορισμός των εργασιών αυτών μπορεί να γίνει για όλο το σύστημα, για κάποια συλλογή, ή για κάποιο τύπο τεκμηρίων. Ένα παράδειγμα εργασίας (για το ΙΑ του ΠΚ) είναι η δημιουργία ενός αντιγράφου ασφαλείας (METS package) για κάθε τεκμήριο που τροποποιείται ενώ είναι δημοσιευμένο. Ένα άλλο είναι η ενημέρωση καταθετών για τη δημοσίευση ή απόρριψη των καταθέσεών τους. Ένα τρίτο, η δημιουργία και αποστολή βεβαίωσης στη γραμματεία της Σχολής ότι δημοσιεύθηκε η (διπλωματική/μεταπτυχιακή/διδακτορική) εργασία ενός φοιτητή. Είναι επίσης δυνατό, νέου τύπου εργασίες να υλοποιηθούν και να χρησιμοποιηθούν δυναμικά (χωρίς ανάπτυξη νέου κώδικα) από το σύστημα.
- **Προσωποποιημένη Διάχυση Περιεχομένου.** Το σύστημα DIAS επιτρέπει σε εγγεγραμμένους χρήστες να δηλώνουν στο προφίλ τους προτιμήσεις για τεκμήρια που τους ενδιαφέρουν με λέξεις κλειδιά. Κατόπιν, στη ροή εργασιών που εκτελούνται κατά τη δημοσίευση ενός τεκμηρίου, μία από τις εργασίες ταυτίζει τα προφίλ χρηστών με τα μεταδεδομένα και το περιεχόμενο του τεκμηρίου και ειδοποιεί τους χρήστες τους οποίους αυτό ενδιαφέρει σύμφωνα με τις προτιμήσεις τους.
- **Διαλειτουργικότητα.** Το σύστημα DIAS είναι πλήρως διαλειτουργικό με τρίτα συστήματα τόσο μέσω της μετάπτωσης και διάθεσης των μεταδεδομένων των τεκμηρίων σε μια σειρά από μορφότυπα (EFRBR[12], HEAL-META[13], OAI-DC[15], METS[14], MODS[4], MARCXML[16]), όσο και μέσω υλοποίησης: α) του πρωτοκόλλου OAI-PMH[17] για άντληση των μεταδεδομένων του (σε διάφορα μορφότυπα), β) του πρωτοκόλλου SWORD[18] για απομακρυσμένη κατάθεση τεκμηρίων από άλλα



συστήματα, και γ) της υπηρεσίας OpenSearch[17] για απομακρυσμένη αναζήτηση περιεχομένου από άλλα συστήματα. Τέλος, παρέχεται και ενσωματώσιμη γραφική διεπαφή (web widget) για την πλοήγηση στις συλλογές και τα τεκμήρια μιας κοινότητας από απομακρυσμένους διαδικτυακούς τόπους (π.χ. τους δικτυακούς τόπους Σχολών, εργαστηρίων κλπ.)

- **Στατιστικά.** Το σύστημα DIAS υποστηρίζει την καταγραφή και παρουσίαση στατιστικών σε επίπεδο συστήματος, συλλογής, αλλά και τεκμηρίου. Τα στοιχεία που διατηρούνται αφορούν σε επισκέψεις, λήψεις, αντλήσεις μεταδεδομένων, αναζητήσεις, κλπ.
- **Αναζήτηση.** Το σύστημα DIAS υποστηρίζει: α) αναζήτηση με λέξεις-κλειδιά (επιτρέποντας και τη χρήση τελεστών (λογικών και ασάφειας) μεταξύ τους) τόσο στα μεταδεδομένα όσο και στο πλήρες κείμενο των τεκμηρίων. β) σύνθετη αναζήτηση με κριτήρια σε συγκεκριμένα γνωρίσματα του σχήματος μεταδεδομένων.
- **Υπηρεσία εξειδικεύσεων του σχήματος μεταδεδομένων.** Το σύστημα DIAS υποστηρίζει τη δημιουργία εξειδικεύσεων (application profiles) στο σχήμα μεταδεδομένων ενός αποθετηρίου δίνοντας τη δυνατότητα ορισμού εξειδικευμένων κανόνων οι οποίοι μπορούν να εφαρμόζονται είτε σε συγκεκριμένες συλλογές, είτε σε συγκεκριμένους τύπους/μορφές τεκμηρίων (π.χ. οι διπλωματικές εργασίες απαιτείται να αναφέρουν τα 3 ως 5 μέλη εξεταστικής επιτροπής, οι μεταπτυχιακές διατριβές του ΗΜΜΥ απαιτείται να αναφέρουν τον τίτλο τους στα Ελληνικά, Αγγλικά, και Ρώσικα. Για το σκοπό αυτό το σύστημα DIAS παρέχει κατάλληλο μηχανισμό δημιουργίας και ελέγχου τέτοιων κανόνων.
- **Πρότυπα Τεκμήρια.** Υποστηρίζεται η δυνατότητα ορισμού προτύπων τεκμηρίων για συγκεκριμένη συλλογή έτσι ώστε ο καταθέτης να έχει προ-συμπληρωμένες τιμές. Για παράδειγμα, οι διπλωματικές εργασίες του Εργαστηρίου TUC/MUSIC έχουν επιβλέπων καθηγητή τον Σταύρο Χριστοδουλάκη, Ακαδημαϊκή Μονάδα το Εργαστήριο Διανεμημένων Πληροφοριακών Συστημάτων, και εκδότη το Πολυτεχνείο Κρήτης.

## Επίλογος

Στο άρθρο αυτό παρουσιάσαμε το σύστημα DIAS ως μια υποδομή δημιουργίας αποθετηρίων. Το σύστημα υιοθετεί το μοντέλο οντοτήτων-συσχετίσεων του προτύπου FRBR για την μοντελοποίηση και καταγραφή τεκμηρίων. Ακολουθεί το παράδειγμα των εγγραφών για την περιγραφή κάθε οντότητας, εισάγοντας όμως σύνθετα γνωρίσματα τα οποία επιτρέπουν την μοντελοποίηση συσχετίσεων τόσο μεταξύ των οντοτήτων του μοντέλου, όσο και με οντότητες έξω από το πλαίσιο του συστήματος. Για παράδειγμα με οντότητες δημοσιευμένες σε αρχεία οντοτήτων (authority files), σε εννοιολογικά συστήματα (concept schemes), σε συστήματα ταξινόμησης (classification schemes), κλπ.

Δείξαμε επίσης, πως οι αρχές μοντελοποίησης του προτύπου SKOS χρησιμοποιήθηκαν από το σύστημα DIAS για τη μοντελοποίηση γνωρισμάτων αναφοράς (συσχέτισης) αλλά και γνωρισμάτων των οποίων η τιμή μπορεί να έχει πολλαπλές (εναλλακτικές) ονομασίες, ή/και μεταφράσεις. Δείξαμε επίσης, πως η χρήση αυτών των μοντελοποιηστικών αρχών ενισχύει την δεικτοδότηση (indexing) και επιλεκτικότητα (selectivity) των τεκμηρίων ενός αποθετηρίου.

Τέλος, παρουσιάσαμε μερικά από τα λειτουργικά χαρακτηριστικά του συστήματος DIAS δείχνοντας κυρίως εκείνα που αυξάνουν την ευελιξία και την προσαρμοστικότητα του συστήματος στις ανάγκες του πραγματικού κόσμου.

## Βιβλιογραφία

- [1] OCLC, Reference Model for an Open Archival Information System (OAIS-RM), <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- [2] LOC, MARC 21 Format for Bibliographic Data, <http://www.loc.gov/marc/bibliographic/>
- [3] IFLA, UNIMARC Core Activity (UCA), <http://www.ifla.org/unimarc>
- [4] LOC, Metadata Object Description Schema (MODS), <http://www.loc.gov/standards/mods/>
- [5] The Dublin Core Metadata Initiative, <http://dublincore.org/>
- [6] Phil Barker, Learning Materials Application Profile (LMAP), <http://www.icbl.hw.ac.uk/lmap/lmapscopingreport.pdf>
- [7] The IEEE 1484.12.1 – 2002 Standard for Learning Object Metadata
- [8] IFLA, Functional Requirements For Bibliographic Records, [https://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf)
- [9] UKOLN: Scholarly Works Application Profile (SWAP), [http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Application\\_Profile](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile)
- [10] W3C, Simple Knowledge Organization System, <https://www.w3.org/2004/02/skos/>
- [11] The Internet Society, Lightweight Directory Access Protocol (LDAP)
- [12] Indiana University, Extended FRBR, <http://purl.org/vocab/frbr/extended>
- [13] ΣΕΑΒ, HEAL-META, [http://orpheus.cn.ntua.gr/schemas/heal/heal\\_v1.2.xsd](http://orpheus.cn.ntua.gr/schemas/heal/heal_v1.2.xsd)
- [14] LOC, Metadata Encoding & Transmission Standard, <http://www.loc.gov/standards/mets/>
- [15] Open Archives Initiative, Dublin Core XML Schema, [http://www.openarchives.org/OAI/2.0/oai\\_dc.xsd](http://www.openarchives.org/OAI/2.0/oai_dc.xsd)
- [16] LOC, MARC21 Xml Schema, <http://www.loc.gov/standards/marcxml/>
- [17] OAI, Protocol for Metadata Harvesting, <https://www.openarchives.org/OAI/openarchivesprotocol.html>
- [18] JISC, Simple Web-service Offering Repository Deposit (SWORD), <http://swordapp.org>
- [19] <http://www.opensearch.org/>