

## 1. Εισαγωγή

Τα ιδρυματικά ηλεκτρονικά αποθετήρια αποτελούν ψηφιακές βάσεις δεδομένων που παρέχουν πρόσβαση μέσω διαδικτύου σε ακαδημαϊκό και ερευνητικό περιεχόμενο. Υλοποιούνται και υποστηρίζονται από ακαδημαϊκούς και ερευνητικούς φορείς, οι οποίοι συγκεντρώνουν, οργανώνουν και διατηρούν σε αυτά ψηφιακό υλικό που αποτελεί προϊόν της λειτουργίας τους και το οποίο διαθέτουν ελεύθερα ή και ελεγχόμενα, αν υπάρχει κάποιος περιορισμός πνευματικών δικαιωμάτων<sup>1</sup>.

Σύμφωνα με σχετική μελέτη των Σπανού, Αρκουλή, Σταύρου, Γιαννοπούλου και Μήτρου (2013) τα οφέλη που μπορεί να προσδώσει η ανάπτυξη και λειτουργία ενός ηλεκτρονικού αποθετηρίου σε ένα ακαδημαϊκό ίδρυμα είναι πολυάριθμα και ιδιαίτερος σημαντικά. Ενδεικτικά αναφέρεται ότι η παρουσία ενός αποθετηρίου μπορεί να:

α) αυξήσει την προβολή και να συνεισφέρει στη διάδοση των ερευνητικών αποτελεσμάτων που παράγονται εντός του ιδρύματος

β) προβάλει το ίδρυμα δίνοντας μια πιο πλήρη εικόνα του σε μελλοντικούς φοιτητές και προσωπικό, αλλά και σε πιθανούς συνεργάτες και ερευνητές

γ) συλλέξει σε ένα κεντρικό σημείο και να συμβάλει στη μακροχρόνια διατήρηση και επικαιροποιημένη αναφορά με το μορφότυπο διάθεση (π.χ. τα αρχεία postScript δύσκολα μπορεί να προσφέρονται σε μορφή pdf) των παραγόμενων ψηφιακών αποτελεσμάτων

δ) συμβάλει στην ποσοτικοποίηση των ερευνητικών και ακαδημαϊκών δραστηριοτήτων και επιδόσεων του ιδρύματος και άρα στην αξιολόγηση της ερευνητικής παραγωγής και αριστείας του συνολικά καθώς και στα επιμέρους επίπεδα της ακαδημαϊκής δομής (παραρτήματα, σχολές, τμήματα, εργαστήρια, ομάδες)

ε) αποτελέσει ένα κεντρικό μέρος συλλογής της τρέχουσας έρευνας που πραγματοποιείται εντός του ιδρύματος

στ) συμβάλει στην προσέλκυση ερευνητικών επιχορηγήσεων και στη σύσταση κοινοπραξιών για τη συμμετοχή σε μεγάλα ερευνητικά έργα

ζ) ενισχύσει τη συνεργασία μεταξύ διαφορετικών ερευνητών Ιδρυμάτων καθώς και τη διεπιστημονική έρευνα

η) συγκεντρώνει και να παρέχει στους φοιτητές διαλέξεις και λοιπό διδακτικό υλικό

θ) υποστηρίζει τη μελέτη και έρευνα των φοιτητών παρέχοντας πρόσβαση σε παλαιότερες

πτυχιακές/διπλωματικές εργασίες και διατριβές.

Στο εξής, χρησιμοποιούνται συχνά οι όροι *τεκμήριο* και *εγγραφή* για την αναφορά στο περιεχόμενο του αποθετηρίου. Αν και η ακριβής χρήση των όρων αυτών θα φαίνεται συνήθως από τα συμφραζόμενα, ο όρος *τεκμήριο* χρησιμοποιείται για να περιγράψει συνολικά ένα έργο (π.χ. μια επιστημονική δημοσίευση, μια πτυχιακή εργασία) ή την έκφραση ενός έργου (π.χ. ένα αρχείο εικόνας που αναπαριστά ένα ζωγραφικό πίνακα) και μπορεί να περιλαμβάνει περισσότερα του ενός αρχεία<sup>2</sup>, ενώ ο όρος *εγγραφή* χρησιμοποιείται για την αναφορά στα μεταδεδομένα ενός έργου. Επίσης, οι όροι *ομάδα τεκμηρίων* και *συλλογή* χρησιμοποιούνται μέσα στο κείμενο, για να δηλώσουν ένα σύνολο τεκμηρίων που έχουν ομαδοποιηθεί με βάση κοινά χαρακτηριστικά τους (π.χ. το σύνολο των πτυχιικών εργασιών ενός τμήματος ενός ιδρύματος) για καλύτερη οργάνωση του περιεχομένου ενός αποθετηρίου.

Προκειμένου να επιτευχθούν τα προαναφερόμενα οφέλη, ένα ακαδημαϊκό ιδρυματικό αποθετήριο χρειάζεται να διαθέτει τη δυνατότητα οργάνωσης και παρουσίασης του περιεχομένου του με ένα δομημένο τρόπο που θα βασίζεται σε κάποιο, πιθανώς μεταβαλλόμενο κατά τη διάρκεια λειτουργίας του, σύστημα ομαδοποίησης. Τα συστήματα ομαδοποίησης που χρησιμοποιούνται κατά κανόνα σε ηλεκτρονικά αποθετήρια είναι ιεραρχικά, και αντιστοιχίζουν το σύνολο των τεκμηρίων σε μια πολυεπίπεδη δομή, δηλαδή συχνά σε περισσότερες από μια διαφορετικές συλλογές (ομάδες ή κλάσεις τεκμηρίων) με βάση ένα ή περισσότερα ποιοτικά κριτήρια που επιλέγονται από το ίδρυμα και εφαρμόζονται από το διαχειριστή του συστήματος. Για παράδειγμα, η ομαδοποίηση των τεκμηρίων μπορεί να ακολουθεί και να αντανάκλα κάποια εξωτερική οργανωτική δομή (π.χ. Σχολές/Τμήματα/Τομείς του ιδρύματος), γεγονός συνηθισμένο ιδιαίτερα σε ιδρύματα με πολυπληθές ανθρώπινο δυναμικό. Σε τέτοιες περιπτώσεις, η ύπαρξη μιας δενδρικής ή γενικότερα, ιεραρχικής δομής διευκολύνει τη διαχείριση, αναζήτηση και εμφάνιση των τεκμηρίων, ενώ βελτιώνει παράλληλα την απόδοση, την ασφάλεια καθώς και τη διατηρησιμότητα των τεκμηρίων στο σύστημα. Τα πλεονεκτήματα από τη δυνατότητα ομαδοποίησης τεκμηρίων σε συλλογές αυξάνονται, αν επιπλέον προσφέρεται και η δυνατότητα ορισμού διαφορετικών επιλογών διαχείρισης και κατάθεσης σε κάθε συλλογή, κατάλληλα προσαρμοσμένων στις απαιτήσεις συγκεκριμένων ομάδων χρηστών του ιδρύματος. Με αυτόν τον τρόπο, το ίδιο σύστημα αποθετηρίου είναι σε θέση να ικανοποιήσει τις διαφορετικές ανάγκες του προσωπικού μεγάλων και επιστημονικά ετερογενών ιδρυμάτων.

Εντούτοις, προκειμένου η δημιουργία των συλλογών τεκμηρίων να είναι δυναμική, σύμφωνη με τις μεταβαλλόμενες απαιτήσεις των χρηστών του αποθετηρίου, αλλά και με το καθαυτό εννοιολογικό περιεχόμενο των τεκμηρίων του (σημαιοσυνολία) το οποίο μόνο a posteriori είναι δυνατόν να αναλυθεί και περιγραφεί, θεωρούμε ότι η ομαδοποίηση των τεκμηρίων σε συλλογές θα πρέπει να βασίζεται σε σχετικά με το περιεχόμενο μεταδεδομένα, που επιπλέον της ολοκλήρωσης περιεχομένου από τα διαφορετικά αποθετήρια θα οδηγήσουν στην επίτευξη της σημασιολογικής διαλειτουργικότητας.

Τα μεταδεδομένα παίζουν πολύ σημαντικό ρόλο στην οργάνωση, περιγραφή και εντοπισμό, συσχέτιση και

<sup>1</sup> Εθνικό Κέντρο Τεκμηρίωσης, «Ηλεκτρονικά Αποθετήρια: Τι Είναι», [http://www.openaccess.gr/repositories/what\\_dot/](http://www.openaccess.gr/repositories/what_dot/) [Ημερομηνία Πρόσβασης 09/12/2012].

<sup>2</sup> Οι όροι έργο, έκφραση έργου και αρχείο είναι σύμφωνοι με το μοντέλο FRBR

ομαδοποίηση και γενικότερα αξιοποίηση των ψηφιακών τεκμηρίων ενός ιδρυματικού αποθετηρίου.

Ανάλογα με τον τύπο τους, τα μεταδεδομένα έχουν και διαφορετική χρησιμότητα:

- τα περιγραφικά μεταδεδομένα αναφέρονται σε ιδιότητες ενός τεκμηρίου, όπως το είδος, ο συγγραφέας ή ο τίτλος του, διευκολύνουν την αναζήτηση και επιταχύνουν την ανακάλυψη του τεκμηρίου από τους χρήστες
- τα δομικά μεταδεδομένα περιγράφουν τη δομή σύνθετων τεκμηρίων που αποτελούνται από περισσότερα του ενός αρχεία (για παράδειγμα, ένα βιβλίο του οποίου κάθε κεφάλαιο είναι αποθηκευμένο σε ξεχωριστό αρχείο) και ως εκ τούτου, βοηθούν να διατηρείται η ακεραιότητα ενός τεκμηρίου ακόμα και όταν τα συστατικά του μέρη είναι αποθηκευμένα σε διαφορετικές τοποθεσίες. Ως δομικά μπορεί να θεωρηθούν επίσης και τα μεταδεδομένα που περιγράφουν τη δομή του συστήματος ομαδοποίησης του αποθετηρίου, καθορίζοντας την ιεραρχία των ομάδων (συλλογών) τεκμηρίων
- τα διαχειριστικά μεταδεδομένα διακρίνονται με τη σειρά τους σε τεχνικά, μεταδεδομένα πρόσβασης και μεταδεδομένα διατήρησης και είναι απαραίτητα για τη διαχείριση των τεκμηρίων εντός του αποθετηρίου. Τα τεχνικά μεταδεδομένα αναφέρονται σε τεχνικές ιδιότητες, όπως το μορφότυπο ενός αρχείου ή η συμμετοχή ενός τεκμηρίου σε μια συλλογή, τα μεταδεδομένα πρόσβασης περιγράφουν τους χρήστες και τις ομάδες χρηστών που έχουν πρόσβαση σε ένα τεκμήριο ή μια συλλογή καθώς και το είδος της πρόσβασης που τους ανατίθεται (για παράδειγμα, δικαίωμα ανάγνωσης, κατάθεσης, επεξεργασίας κλπ), ενώ τα μεταδεδομένα διατήρησης περιλαμβάνουν τις απαραίτητες πληροφορίες για τη μακρόχρονη διατήρηση ενός τεκμηρίου, όπως για παράδειγμα η ακριβής τοποθεσία αποθήκευσης και τα αθροίσματα ελέγχου των συστατικών μερών του.

Πολλά είναι τα σχήματα μεταδεδομένων που χρησιμοποιούνται για την επισημείωση των τεκμηρίων που μπορεί να περιέχονται σε ένα ακαδημαϊκό ιδρυματικό αποθετήριο με συχνότερα χρησιμοποιούμενα τα: Dublin Core (DC) / ETD-MS / SWAP / XMetaDiss για πτυχιακές εργασίες και διπλωματικές διατριβές, IEEE LOM για εκπαιδευτικό υλικό, PRISM για δημοσιεύσεις σε περιοδικά κ.ά. Πρόσφατα, προτάθηκε από το Σύνδεσμο Ελληνικών Ακαδημαϊκών Βιβλιοθηκών – ΣΕΑΒ (Σπανός, Αρκουλή, Σταύρου, Γιαννοπούλου και Μήτρου 2013) η υιοθέτηση του σχήματος μεταδεδομένων healMeta από το σύνολο των ελληνικών ακαδημαϊκών ιδρυμάτων ως ειδικού για την περιγραφή της ακαδημαϊκής και ερευνητικής παραγωγής, το οποίο δεν αποκλείει τη συνύπαρξη με DC, ενισχύει τη σημασιολογική διαλειτουργικότητα μεταξύ αποθετηρίων που χρησιμοποιούν διαφορετικές τακτικές συμπλήρωσης μεταδεδομένων και βοηθά στη συγκομιδή μεταδεδομένων μέσω OAI-PMH από κατάλληλους συγκεντρωτές περιεχομένου.

Στην παρούσα εργασία περιγράφεται η επέκταση του healMeta με προσθήκη κατάλληλων στοιχείων για τη δημιουργία συλλογών στο ιδρυματικό αποθετήριο του ΤΕΙ Κρήτης, το οποίο περιέχει τεκμήρια σχετικά με τις πτυχιακές εργασίες των φοιτητών. Το όνομα του επεκτεταμένου σχήματος, TCDMeta, αντιστοιχεί στο είδος των τεκμηρίων στα μεταδεδομένα των οποίων αναφέρεται: TEI of Crete Diploma theses, ενώ η διάρθρωση του εγγράφου περιλαμβάνει στην ενότητα 2 τη μεθοδολογία που ακολουθήθηκε για τον προσδιορισμό βάσει εννοιολογικής ανάλυσης των συλλογών που πρέπει να δημιουργηθούν στο αποθετήριο, στην ενότητα 3 τον ορισμό των μεταδεδομένων που σχετίζονται με τη δημιουργία νέων συλλογών και στην ενότητα 4 συμπεράσματα και μελλοντικές επεκτάσεις του προτεινόμενου σχήματος προκειμένου να χρησιμοποιηθεί και σε άλλες κατηγορίες τεκμηρίων.

## **2. Μεθοδολογία για τον καθορισμό νέων συλλογών στο ιδρυματικό αποθετήριο του ΤΕΙ Κρήτης**

Στο ΤΕΙ Κρήτης, έχει ξεκινήσει από το 2001 μία προσπάθεια που αφορά στη συλλογή, διατήρηση και προβολή, με ενιαίο τρόπο, του πνευματικού έργου που παράγεται κυρίως από τους φοιτητές στο ΤΕΙ και δε δημοσιεύεται αλλού. Το υλικό που περιλαμβάνει αποτελείται από πτυχιακές εργασίες των φοιτητών όλων των Τμημάτων του ΤΕΙ. Στο ιδρυματικό αποθετήριο του ΤΕΙ Κρήτης (e-Thesis: <http://nefeli.lib.teicrete.gr>) μπορεί κάθε ενδιαφερόμενος να έχει πρόσβαση στα πλήρη κείμενα των πτυχιακών εργασιών, η περιγραφή των οποίων γίνεται με ένα απλό «βιβλιογραφικό σχήμα» που δανείζεται τα στοιχεία του από το DC, χωρίς εντούτοις να το χρησιμοποιεί καθεαυτό.

Στο παρόν ιδρυματικό αποθετήριο η ομαδοποίηση των τεκμηρίων είναι σχετικά απλή και γίνεται ανά Σχολή, Τμήμα και έτος δημιουργίας. Επιπλέον, ο χρήστης έχει τη δυνατότητα να κάνει αναζήτηση με ελεύθερο κείμενο στις βιβλιογραφικές περιγραφές ή σε επιμέρους πεδία περιγραφικών μεταδεδομένων των τεκμηρίων (τίτλος, συγγραφέας, επιβλέπωντας, περίληψη, έτος).

Αν και απλή, η ομαδοποίηση των τεκμηρίων είναι σημαντική και δίνει στο χρήστη τη δυνατότητα να δει επιμέρους υποσυλλογές του συστήματος, όμως δεν θεωρείται αρκετή για την έρευνα και την πιο ουσιαστική χρήση του περιεχομένου του αποθετηρίου. Θεωρήθηκε γι' αυτό σημαντική και μελετήθηκε η δυνατότητα περαιτέρω ομαδοποιήσεων με βάση άλλες ιεραρχικές δομές ώστε να διευκολύνεται η αναζήτηση και εμφάνιση των τεκμηρίων. Μάλιστα αυτό είναι και ένα από τα ζητούμενα της διαδικασίας αναβάθμισης της υπηρεσίας του ιδρυματικού αποθετηρίου του ΤΕΙ Κρήτης, η οποία υλοποιείται στο πλαίσιο της πράξης «Οργάνωση, ανάδειξη και προβολή του ακαδημαϊκού περιεχομένου του ΤΕΙ Κρήτης» του επιχειρησιακού προγράμματος «Ψηφιακή σύγκλιση».

Ο τρόπος δημιουργίας νέων συλλογών και μετα-συλλογών τεκμηρίων (θεματικών και μη) καθορίστηκε κατά τη διάρκεια εκσφαλμάτωσης των περιγραφικών μεταδεδομένων των τεκμηρίων του υπάρχοντος ιδρυματικού αποθετηρίου προκειμένου αυτά να εισαχθούν με τη βέλτιστη ποιότητα και πληρότητα στο νέο σύστημα που πρόκειται να υλοποιηθεί. Έτσι χρειάστηκε να μελετηθεί συνολικά η συλλογή 5157 πτυχιακών εργασιών και των σχετικών με αυτές τεκμηρίων και εγγραφών και να πραγματοποιηθούν διορθώσεις στο σύνολό της.

Κατά τη μελέτη και διόρθωση του περιεχομένου του αποθετηρίου επιπλέον διαπιστώθηκε η ανάγκη να καταρτιστεί λίστα με προτεινόμενους όρους που θα χρησιμοποιούνται ως λέξεις κλειδιά στη θέση αυτών που έχουν ατυχώς επιλεγεί. Αναφορικά με αυτή τη λίστα, προέκυψε τελικά ένας ενιαίος αλφαβητικός κατάλογος με τις λέξεις

κλειδιά που έχουν χρησιμοποιηθεί στο υπάρχον σύστημα. Κατά την επεξεργασία των λέξεων κλειδιών παρατηρήθηκε το φαινόμενο της συχνής επανάληψης συγκεκριμένων όρων, ενώ από τη μελέτη της συλλογής παρατηρήθηκε η επανάληψη συγκεκριμένων τίτλων πτυχιικών εργασιών (ίδιο θέμα), η εκπόνηση εργασιών από διαφορετικά επιστημονικά πεδία, που όμως αλληλεπιδρούν κ.ά.

Στη συνέχεια παρουσιάζονται αναλυτικότερα οι νέες συλλογές τεκμηρίων που μπορεί να προκύψουν από το ήδη καταχωρημένο περιεχόμενο στο σύστημα:

- *Δημοφιλή θέματα εργασιών ή θεματικές συλλογές* σχετικές με την επαναληψιμότητα των λέξεων κλειδιών. Εξετάζοντας το πόσες φορές εμφανίζεται στο σύστημα μία λέξη κλειδί, και σε ποιες εργασίες, μπορεί κανείς να εξάγει κάποια δημοφιλή θέματα εργασιών. Θα μπορούσε εντούτοις να δημιουργηθούν θεματικές συλλογές με βάση τις λέξεις κλειδιά που έχουν καταχωρηθεί σε εγγραφές. Να σημειωθεί ότι στο υπάρχον ιδρυματικό αποθετήριο οι λέξεις κλειδιά που χρησιμοποιούνται για τον χαρακτηρισμό της κάθε εργασίας δεν προήλθαν εξαρχής από κάποιο ελεγχόμενο λεξιλόγιο όρων ή θησαυρό. Ήταν όροι που δόθηκαν ή από τον συγγραφέα της εργασίας, είτε από τον υπεύθυνο για την καταχώρησή της. Το γεγονός αυτό είχε σαν αποτέλεσμα να δίδονται ίδιες λέξεις κλειδιά με πολλούς διαφορετικούς τρόπους (π.χ. συντομογραφίες, χρήση ενικού ή πληθυντικού, ακόμα και ορθογραφικά λάθη). Με τη δημιουργία του αλφαβητικού καταλόγου προτεινόμενων όρων για τις λέξεις κλειδιά, τα προαναφερθέντα προβλήματα εξαλείφθηκαν. Έτσι η δημιουργία θεματικών συλλογών από τις λέξεις κλειδιά έγινε αμεσότερη και ορθή. Επιπλέον, το σύστημα που πρόκειται να υλοποιεί μελλοντικά το ιδρυματικό αποθετήριο θα περιλαμβάνει σε κατάλληλο στάδιο της ροής καταχώρησης εργασιών τη δυνατότητα καταχώρησης της θεματικής κατηγορίας στην οποία καθεμιά αντιστοιχεί από τις θεματικές επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου (Library of Congress Subject Headings), τους καθιερωμένους όρους της Εθνικής Βιβλιοθήκης, ή θεματικές επικεφαλίδες. Αυτή η λειτουργικότητα θα βοηθήσει ακόμη περισσότερο τη δημιουργία θεματικών συλλογών σχετικών πλέον με τις χρησιμοποιούμενες θεματικές επικεφαλίδες.
- *Συλλογές πτυχιικών εργασιών με θέμα που επαναλαμβάνεται* οι οποίες μπορεί να προκύψουν από τους τίτλους των εργασιών. Κατά τη μελέτη της συλλογής παρατηρήθηκε η ύπαρξη εργασιών με ίδιο ή παρόμοιο τίτλο. Αν και αρχικά θεωρήθηκε ως αστοχία των εκπαιδευτικών που προτείνουν τα θέματα και επιβλέπουν τις εργασίες, με δεύτερη ματιά σκεφτήκαμε ότι η δημιουργία τέτοιων συλλογών θα συμβάλει στη μακροσκοπική μελέτη των αποτελεσμάτων των εργασιών. Έτσι, ο μελλοντικός ερευνητής θα μπορεί λόγω χάρη να εξετάσει την παραγωγικότητα ή ενδεχόμενη παραγωγικότητα σε βάθος χρόνου από την εφαρμογή διαφορετικών μεθόδων ή πειραμάτων σε συγκεκριμένα επιστημονικά πεδία, να δει την εξέλιξη των συστημάτων που χρησιμοποιούνται σε ίδια πεδία εφαρμογής, ή ακόμη τις επιπτώσεις της αλλαγής περιβάλλοντος στην εφαρμογή επιστημονικών λύσεων. Πρόκειται για συλλογές – μεταεργασίες, που μπορεί να έχουν το ρόλο των «παρατηρητηρίων επιστήμης και τεχνολογίας».
- *Συλλογές με βάση γεωγραφικές περιοχές*. Αυτές μπορεί να περιλαμβάνουν εργασίες που αφορούν στην πραγματοποίηση πειραμάτων, ερευνών και μεθόδων, ή την υλοποίηση συστημάτων και υπηρεσιών σε μία συγκεκριμένη γεωγραφική περιοχή. Χρησιμοποιώντας τα τοπωνύμια που αναφέρονται στον τίτλο, την περίληψη και τις λέξεις κλειδιά, μπορούν να δημιουργηθούν συλλογές για τις γεωγραφικές περιοχές. Για παράδειγμα, θα μπορεί ο ερευνητής να δει τι εργασίες έχουν γίνει για την Κρήτη, Ελλάδα, Ευρώπη κλπ. Στις συγκεκριμένες γεωγραφικές συλλογές θα μπορούσε να δημιουργηθούν υποσυλλογές σχετικές με το τμήμα στο οποίο πραγματοποιήθηκε μια εργασία ή τον επιστημονικό τομέα που καλύπτει. Για παράδειγμα, θα μπορεί ο ερευνητής να δει τις μελέτες που έχουν πραγματοποιηθεί για την μόλυνση των υδάτων σε συγκεκριμένες γεωγραφικές περιοχές ή την καλλιέργεια συγκεκριμένων λαχανικών σε μία γεωγραφική περιοχή.
- *Συλλογές εργασιών επιστημονικών τομέων* των τμημάτων των σχολών του ιδρύματος. Είναι πολύ χρήσιμο να μπορεί το σύστημα να δίνει συλλογές με τις πτυχιικές ή μεταπτυχιικές εργασίες που έχουν γίνει σε κάθε τομέα, ή ακόμα και σε κάθε εργαστήριο ενός τμήματος, ιδιαίτερα για την προβολή του έργου του και την εξαγωγή δεικτών αξιολόγησης. Η δημιουργία αυτών των συλλογών μπορεί να γίνει με αντιστοίχιση του επιβλέποντα καθηγητή της εργασίας με τον τομέα ή το εργαστήριο στο οποίο ανήκει.
- *Συλλογή διακεκριμένων πτυχιικών εργασιών*. Για τη δημιουργία της θα πρέπει να οριστούν συγκεκριμένα κριτήρια από τους επιβλέποντες καθηγητές, τα οποία όταν πληρούνται, η εργασία να χαρακτηρίζεται διακεκριμένη και να συμπληρώνεται αυτόματα κατάλληλο στοιχείο του σχήματος μεταδεδομένων. Η συλλογή αυτή μπορεί να είναι ιδιαίτερα σημαντική για την αναγνωρισιμότητα των δημιουργών των εργασιών που περιλαμβάνει (λχ συγκριτικό πλεονέκτημα κατά την επιλογή τους σε μεταπτυχιακά προγράμματα ή θέσεις εργασίας).
- *Συλλογές δημοσιευμένων εργασιών* για διάφορους επιστημονικούς τομείς, όπου θα συμπεριληφθούν πτυχιικές εργασίες που μέρος τους έχει δημοσιευτεί σε κάποιο επιστημονικό περιοδικό ή έχει παρουσιαστεί σε κάποιο συνέδριο. Θα συμπληρώνεται κατά την απόθεση ή σε επόμενη επεξεργασία, κατάλληλο στοιχείο του σχήματος μεταδεδομένων.
- *Συλλογές συγκριτικών μελετών* για διάφορους επιστημονικούς τομείς. Στις συγκεκριμένες συλλογές μπορεί να ομαδοποιηθούν εργασίες που περιέχουν στον τίτλο ή την περίληψη λέξεις όπως έρευνα, συγκριτική μελέτη, μελέτη ή μεταφράσεις των λέξεων αυτών στην αγγλική.
- *Συλλογές εργασιών με ίδιο ή ανάλογο μορφότυπο* των μερών του συνοδευτικού τους υλικού. Σε αυτές θα περιλαμβάνονται λόγω χάρη όλες οι εργασίες που συνοδεύονται από σχέδια σε autocad, ή video σε mp4 και

θα διευκολύνουν όχι μόνο τον ενδιαφερόμενο για τέτοιο υλικό χρήστη αλλά και το διαχειριστή του αποθετηρίου ο οποίος για λόγους διατήρησης προτίθεται να αλλάξει συγκεκριμένους μορφότευπους αρχείων σε επικαιροποιημένες εκδόσεις τους.

Εύκολα μπορεί κανείς να συμπεράνει ότι η αυτοματοποιημένη δημιουργία των συλλογών που προαναφέρθηκαν μπορεί να υποστηριχτεί ουσιαστικά μόνο μέσω της συνεπούς επισημείωσης των τεκμηρίων με κατάλληλα μεταδεδομένα, άρα την υιοθέτηση και χρήση σχήματος που τα περιλαμβάνει. Στην ενότητα που ακολουθεί παρουσιάζονται τα στοιχεία του healMeta που χρησιμοποιούνται για το σκοπό αυτό, καθώς και επιπλέον στοιχεία που δεν περιλαμβάνονται και επομένως μας οδήγησαν στον εμπλουτισμό του απ' όπου προέκυψε το σχήμα μεταδεδομένων TCDMeta.

### 3. Στοιχεία των σχημάτων healMeta και TCDMeta που απαιτούνται για την αυτοματοποιημένη δημιουργία των νέων συλλογών

Αναλυτική παρουσίαση της αναγκαιότητας αλλά και των στοιχείων που περιλαμβάνει το σχήμα μεταδεδομένων healMeta γίνεται σε σχετική μελέτη των Σπανού, Αρκουλή, Σταύρου, Γιαννοπούλου και Μήτρου (2013). Από τα στοιχεία που περιλαμβάνει στη δημιουργία των συλλογών που αναφέρθηκαν στην ενότητα 2 απαραίτητα είναι τα ακόλουθα:

#### Τύπος Έργου

Στοιχείο: heal:type

Πιθανές τιμές: (bachelorThesis, masterThesis, doctoralThesis, conferenceItem, journalArticle, bookChapter, book, report, learningMaterial, dataset, other)

Υποχρεωτικό: Ναι

Επαναληψιμότητα στοιχείου: Όχι

Περιγραφή περιεχομένου: Ο τύπος του έργου, ο οποίος εκφράζεται ως μια τιμή προερχόμενη από ένα προκαθορισμένο λεξιλόγιο όρων. Οι δυνατές τιμές που μπορεί να λάβει το στοιχείο είναι "bachelorThesis" για προπτυχιακή εργασία, "masterThesis" για μεταπτυχιακή εργασία, "doctoralThesis" για διδακτορική διατριβή, "conferenceItem" για δημοσίευση σε συνέδριο, "journalArticle" για περιοδική δημοσίευση, "bookChapter" για κεφάλαιο βιβλίου, "book" για βιβλίο ή μονογραφία, "report" για τεχνική αναφορά, "learningMaterial" για εκπαιδευτικό υλικό, "dataset" για σύνολο δεδομένων και "other" για άλλο είδος έργου.

XML Σύνταξη: <heal:type>Είδος Έργου</heal:type>

Παράδειγμα χρήσης: <heal:type>bachelorThesis</heal:type>

#### Τίτλος

Στοιχείο: heal:title

Γνώρισμα: xml:lang= "κωδικός από RFC 5646" (υποχρεωτικό)

Υποχρεωτικό: Ναι

Επαναληψιμότητα στοιχείου: Ναι

Περιγραφή περιεχομένου: Ο κύριος τίτλος του έργου όπως αυτός έχει αποδοθεί από το δημιουργό του. Το γνώρισμα xml:lang δηλώνει τη γλώσσα στην οποία εκφράζεται ο τίτλος του έργου και δηλώνεται με αντίστοιχο κωδικό που συμφωνεί με το πρότυπο RFC 5646<sup>3</sup> και ο οποίος μπορεί να αναζητηθεί στον αντίστοιχο κατάλογο της IANA<sup>4</sup> (π.χ. "en" για την αγγλική και "el" για την ελληνική γλώσσα).

XML Σύνταξη: <heal:title xml:lang="κωδικός από RFC 5646">Τίτλος Έργου</heal:title>

Παράδειγμα χρήσης: <heal:title xml:lang="el">Διερεύνηση συμπεριφοράς οδηγών ως προς την ταχύτητα κατά την προσέγγιση σε μη σηματοδοτημένο κόμβο</heal:title>

#### Θεματική Κατηγορία

Στοιχείο: heal:classification

Γνώρισμα: scheme= (LCC, DDC, UDC, NLM, ACMCCS, MSC, PACS, other) (προαιρετικό)

Γνώρισμα: xml:lang= "κωδικός από RFC 5646" (υποχρεωτικό)

Υποχρεωτικό: Όχι

Επαναληψιμότητα στοιχείου: Ναι

Περιγραφή περιεχομένου: Η ευρύτερη θεματική κατηγορία στην οποία υπάγεται το έργο, η οποία μπορεί να αποτελεί καθιερωμένο όρο προερχόμενο από κάποιο σχήμα κωδικοποίησης (encoding scheme) ή ελεύθερο κείμενο. Στην πρώτη περίπτωση, το στοιχείο πρέπει να περιλαμβάνει το γνώρισμα scheme που δηλώνει το σχήμα κωδικοποίησης που χρησιμοποιείται για τη δήλωση της θεματικής κατηγορίας. Οι πιθανές τιμές του γνωρίσματος scheme είναι "LCC" για το Library of Congress Classification, "DDC" για το Dewey Decimal Classification System, "UDC" για το Universal Decimal Classification System, "NLM" για το National Library of Medicine Classification, "ACMCCS" για το ACM Computing Classification System, "MSC" για το Mathematics Subject Classification και "PACS" για το Physics and Astronomy Classification Scheme. Για λόγους σημασιολογικής διαλειτουργικότητας μεταξύ των ελληνικών ΙΑ, προτείνεται η χρήση των συστημάτων ταξινόμησης LCC και DDC, για τα οποία έχουν προταθεί και σχετικοί πίνακες αντιστοίχισης. Η τιμή του στοιχείου είναι ο τίτλος της θεματικής κατηγορίας, ακολουθούμενος από τον αντίστοιχο κωδικό που ορίζει το σχήμα κωδικοποίησης εντός

<sup>3</sup> <http://tools.ietf.org/html/rfc5646>

<sup>4</sup> IANA Language Subtag Registry: <http://www.iana.org/assignments/language-subtag-registry>



αγκυλών, δηλαδή έχει την μορφή: «*Τίτλος θεματικής κατηγορίας*» [*«Κωδικός θεματικής κατηγορίας»*]. Στην περίπτωση που για τη δήλωση της θεματικής κατηγορίας, χρησιμοποιείται κάποιος ελεύθερος όρος, το γνώρισμα scheme παραλείπεται. Και στις δύο περιπτώσεις, το γνώρισμα xml:lang δηλώνει τη γλώσσα στην οποία εκφράζεται η θεματική κατηγορία. Αξίζει να σημειωθεί ότι το στοιχείο **Θεματική Κατηγορία** περιγράφει μόνο το ευρύτερο γνωστικό θεματικό αντικείμενο του έργου, σε αντίθεση με το στοιχείο **Λέξη-κλειδί**, που περιέχει περισσότερο ειδικούς όρους.

*XML Σύνταξη:* <heal:classification xml:lang="κωδικός από RFC 5646" scheme="αναγνωριστικό σχήματος κωδικοποίησης">Θεματική κατηγορία [Κωδικός θεματικής κατηγορίας]</heal:classification>

*Παράδειγμα χρήσης:* <heal:classification xml:lang="en" scheme="LCC">Surveying engineering science [Q161]</heal:classification>

<heal:classification xml:lang="el">Επιστήμη τοπογράφου μηχανικού</heal:classification>

#### **Λέξη-κλειδί (Θέμα)**

*Στοιχείο:* heal:keyword

*Γνώρισμα:* scheme= (LCSH, MeSH, STW, AAT, other) (προαιρετικό)

*Γνώρισμα:* xml:lang= "κωδικός από RFC 5646" (υποχρεωτικό)

*Υποχρεωτικό:* Όχι

*Επαναληψιμότητα στοιχείου:* Ναι

*Περιγραφή περιεχομένου:* Μια λέξη-κλειδί που περιγράφει το έργο, η οποία μπορεί να αποτελεί καθιερωμένο όρο προερχόμενο από κάποιο θησαυρό ή λεξιλόγιο όρων ή ελεύθερο κείμενο. Στην πρώτη περίπτωση, το στοιχείο πρέπει να περιλαμβάνει το γνώρισμα scheme που δηλώνει το λεξιλόγιο από το οποίο προέρχεται ο θεματικός όρος. Οι πιθανές τιμές του γνωρίσματος scheme είναι "LCSH" για τα Library of Congress Subject Headings, "MeSH" για τα Medical Subject Headings, "STW" για τον ομώνυμο θησαυρό οικονομικών όρων και "AAT" για τον Getty Art & Architecture Thesaurus. Η τιμή του στοιχείου είναι η λέξη-κλειδί, ακολουθούμενη από τον αντίστοιχο κωδικό που ορίζει το λεξιλόγιο όρων εντός αγκυλών, δηλαδή έχει την μορφή: «*Λέξη-κλειδί*» [*«Κωδικός λέξης»*]. Στην περίπτωση που η λέξη-κλειδί δεν προέρχεται από κάποιο λεξιλόγιο, το γνώρισμα scheme παραλείπεται. Και στις δύο περιπτώσεις, το γνώρισμα xml:lang δηλώνει τη γλώσσα στην οποία εκφράζεται η θεματική κατηγορία.

*XML Σύνταξη:* <heal:keyword xml:lang="κωδικός από RFC 5646" scheme="αναγνωριστικό λεξιλογίου όρων">Λέξη-κλειδί [Κωδικός λέξης]</heal:keyword>

*Παράδειγμα χρήσης:* <heal:keyword xml:lang="en" scheme="LCSH">Linear models (Statistics) [sh85077177]</heal:keyword>

<heal:keyword xml:lang="el">Μη σηματοδοτημένος κόμβος</heal:keyword>

#### **Μορφότυπο περιεχόμενου αρχείου**

*Στοιχείο:* heal:fileFormat

*Υποχρεωτικό:* Όχι

*Επαναληψιμότητα στοιχείου:* Ναι

*Περιγραφή περιεχομένου:* Το μορφότυπο των αρχείων που αναπαριστούν το έργο, εκφρασμένο σύμφωνα με το λεξιλόγιο μορφότυπων του IANA<sup>5</sup>. Η παρουσία του εν λόγω στοιχείου είναι προαιρετική, για να καλυφθεί το ενδεχόμενο ύπαρξης εγγραφών χωρίς συνοδευτικά αρχεία.

*XML Σύνταξη:* <heal:fileFormat>Μορφότυπο αρχείου</heal:fileFormat>

*Παράδειγμα χρήσης:* <heal:fileFormat>application/pdf</heal:fileFormat>

<heal:fileFormat>application/vnd.ms-powerpoint</heal:fileFormat>

#### **Όνομα Επιβλέποντος**

*Στοιχείο:* heal:advisorName

*Γνώρισμα:* xml:lang= "κωδικός από RFC 5646" (υποχρεωτικό)

*Υποχρεωτικό:* Ναι

*Επαναληψιμότητα στοιχείου:* Όχι

*Περιγραφή περιεχομένου:* Το όνομα του επιβλέποντος μιας φοιτητικής εργασίας. Η δομή και το περιεχόμενο αυτού του στοιχείου ακολουθούν την ίδια λογική με το στοιχείο **Όνομα Δημιουργού**.

*XML Σύνταξη:* <heal:advisorName xml:lang="κωδικός από RFC 5646">Όνομα επιβλέποντος</heal:advisorName>

*Παράδειγμα χρήσης:* <heal:advisorName xml:lang="el">Αντωνίου, Κωνσταντίνος</heal:advisorName>

Επιπλέον των παραπάνω απαιτούνται στοιχεία για το χαρακτηρισμό εργασιών ως διακεκριμένες ή ως συγκριτικές μελέτες. Τα στοιχεία αυτά που περιλαμβάνονται στο σχήμα μεταδεδομένων TCDMeta ορίζονται ως εξής:

#### **Χαρακτηρισμός ως Διακεκριμένη**

*Στοιχείο:* TCD:Distinguished

*Υποχρεωτικό:* Ναι

*Επαναληψιμότητα στοιχείου:* Όχι

*Περιγραφή περιεχομένου:* Ένδειξη για το αν η εργασία είναι διακεκριμένη. Ο χαρακτηρισμός δηλώνεται με την τιμή true και η μη χαρακτηρισμός με την τιμή false, υπακούοντας στην προδιαγραφή του τύπου δεδομένων

<sup>5</sup> IANA list of MIME Media Types: <http://www.iana.org/assignments/media-types/index.html>

boolean του XML Schema.

XML Σύνταξη: `<TCD:Distinguished> {true, false} </ TCD:Distinguished >`

Παράδειγμα χρήσης: `< TCD:Distinguished > true </ TCD:Distinguished >`

#### Χαρακτηρισμός ως Συγκριτική Μελέτη

Στοιχείο: TCD:Survey

Υποχρεωτικό: Ναι

Επαναληψιμότητα στοιχείου: Όχι

Περιγραφή περιεχομένου: Ένδειξη για το αν η εργασία αποτελεί συγκριτική μελέτη ή όχι. Ο χαρακτηρισμός δηλώνεται με την τιμή true και η μη χαρακτηρισμός με την τιμή false, υπακούοντας στην προδιαγραφή του τύπου δεδομένων boolean του XML Schema.

XML Σύνταξη: `<TCD:Survey> {true, false} </ TCD: Survey>`

Παράδειγμα χρήσης: `< TCD: Survey> false </ TCD: Survey>`

Είναι φανερό, ότι η διαφοροποίηση των δυο σχημάτων μεταδομένων έγκειται μόνο στα στοιχεία εκείνα που σχετίζονται με τη δημιουργία συλλογών διακεκριμένων εργασιών και συγκριτικών μελετών. Εντούτοις εάν προκύψει αναγκαιότητα να δημιουργηθεί οποιαδήποτε άλλη συλλογή (είτε από περαιτέρω ανάλυση του περιεχομένου του αποθετηρίου, ή από εννοιολογικά χαρακτηριστικά καινούργιου περιεχομένου που θα προστεθεί σε αυτό), διαπιστώνει κανείς ότι είναι εύκολο να γίνει με προσθήκη κατάλληλων στοιχείων στο σχήμα healMeta.

#### 4. Συμπεράσματα – μελλοντικές επεκτάσεις

Βασικό συμπέρασμα της εργασίας αυτής αποτελεί το γεγονός ότι για την επίτευξη της μέγιστης δυνατής ωφέλειας από τα τεκμήρια που περιλαμβάνονται σε ένα ακαδημαϊκό αποθετήριο, τόσο για το ίδρυμα όσο και για τους δημιουργούς και τους τελικούς χρήστες – μελετητές, σημαντική είναι η δημιουργία συλλογών τεκμηρίων με χρήση σχημάτων μεταδεδομένων καθώς:

α) αυτά εύκολα και ελεγχόμενα μπορεί να επεκτείνονται και χρησιμοποιούνται ώστε να επιτρέπουν την σημασιολογική επισημείωση των τεκμηρίων, και

β) η χρήση τους διευκολύνει την αυτοματοποίηση της δημιουργίας των απαιτούμενων συλλογών.

Επιπλέον, διαπιστώσαμε την αξία της a posteriori μελέτης μιας συλλογής τεκμηρίων για τη λήψη αποφάσεων σχετικών με τη βέλτιστη ομαδοποίησή τους, καθώς όσο καλή κι αν είναι μια a priori μελέτη, παραμένει εκτίμηση όσο τα δεδομένα (τεκμήρια) έχουν (και κατά τη γνώμη μας θα έχουν πάντα) τη δική τους δυναμική και σημασία.

Σχεδιάζουμε και πρόκειται να υλοποιήσουμε την υποστήριξη του σχήματος TCDMeta (και επομένως του healMeta) στο νέο σύστημα ιδρυματικού αποθετηρίου του ΤΕΙ Κρήτης, με τη βοήθεια των οποίων θα αυτοματοποιήσουμε τη δημιουργία των συλλογών στις οποίες αναφερθήκαμε.

#### Αναφορές

- Allard, S., Mack, T. R., Felther-Reichert, M. (2005), "The librarian's role in institutional repositories: A content analysis of the literature", *Reference Services Review*, Vol.33 Iss: 3 pp.325-336.
- Burk, A., Al-Digeil, M., Forest, D., Whitney, J. (2007) "New possibilities for metadata creation in an institutional repository context", *OCLC Systems & Services*, Vol. 23 Iss: 4, pp.403 – 410
- Jain, P., (2011), "New trends and future applications/directions of institutional repositories in academic institutions", *Library Review*, Vol.60 Iss: 2 pp.125-141.
- International Federation of Library Associations and Institutions (IFLA), Functional Requirements for Bibliographic Records, 1997.
- Lagoze C., van de Sompel H., Nelson M., Warner S., "The Open Archives Initiative Protocol for Metadata Harvesting", 2008, available online at: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
- M. Nilsson, A. Powell, P. Johnston, and A. Naeve, "Expressing Dublin Core metadata using the Resource Description Framework (RDF)", 2008, available online at: <http://www.dublincore.org/documents/dc-rdf>.
- Μορέλη-Κακούρη Μ., Μακρίδου Ε., Ασμανίδης Υ. (2007): Ιδρυματικά Καταθετήρια: Προτάσεις για μια Ενιαία Εθνική Πολιτική με Βάση ένα Ελληνικό Παράδειγμα. *Πρακτικά του 16ου Πανελληνίου Συνεδρίου Ακαδημαϊκών Βιβλιοθηκών: Ο Ανθρώπινος Παράγοντας στη Διαμόρφωση της Σημερινής και της Μελλοντικής Βιβλιοθήκης*, 494 – 512.
- Σφακιανάκης Μ., Καπιδάκης Σ. (2007): Ενισχύοντας Σημασιολογικά τις Διαδικασίες Αναζήτησης σε Ένα Περιβάλλον Μετα-Αναζήτησης. *Πρακτικά του 16ου Πανελληνίου Συνεδρίου Ακαδημαϊκών Βιβλιοθηκών: Ο Ανθρώπινος Παράγοντας στη Διαμόρφωση της Σημερινής και της Μελλοντικής Βιβλιοθήκης*, 640 – 651.
- Σταυράκης, Ν., Τοράκη, Κ. (2005): Συστήματα Διαχείρισης Ηλεκτρονικών Αρχείων: Παρουσίαση και Συγκριτική Αξιολόγηση Τριών Μοντέλων Μεταδεδομένων. *Πρακτικά του 14ου Πανελληνίου Συνεδρίου Ακαδημαϊκών Βιβλιοθηκών: Διαχείριση της Γνώσης: Ο Παγκόσμιος Ιστός και οι Μονάδες Πληροφόρησης*, 325 – 334.
- Δ. Σπανός, Σ. Αρκουλής, Π. Σταύρου, Ε. Γιαννοπούλου και Ν. Μήτρου, «ΛΕΙΤΟΥΡΓΙΚΕΣ ΚΑΙ ΤΕΧΝΙΚΕΣ ΠΡΟΔΙΑΓΡΑΦΕΣ ΙΔΡΥΜΑΤΙΚΩΝ ΑΠΟΘΕΤΗΡΙΩΝ» Έκδοση 1.1, 2013, διαθέσιμο μέσω του: [http://seab.lib.ntua.gr/index.php?option=com\\_docman&task=doc\\_download&gid=12&Itemid=53&lang=el](http://seab.lib.ntua.gr/index.php?option=com_docman&task=doc_download&gid=12&Itemid=53&lang=el)
- Εθνικό Κέντρο Τεκμηρίωσης, «Ηλεκτρονικά Αποθετήρια: Τι Είναι», [http://www.openaccess.gr/repositories/what\\_dot/](http://www.openaccess.gr/repositories/what_dot/).
- Dublin Core Collections Application Profile, 2007, available online at: <http://dublincore.org/groups/collections/collection-application-profile/>.
- ETD-MS standard: <http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>
- Scholarly Works Application Profile, [http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Application\\_Profile](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile)
- XMetaDiss format: <http://www.d-nb.de/eng/standards/xmetadiss/xmetadiss.htm>
- Final 1484.12.1-2002 LOM Draft Standard: <http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html>

PRISM Specification: <http://www.idealliance.org/specifications/prism/>  
DCMI Metadata Terms: <http://dublincore.org/documents/dcmi-terms/>