

Επιτυχημένα παραδείγματα διαλειτουργικότητας σε ελληνικά αποθετήρια και σχετικά τεχνολογικά εργαλεία

Νίκος Κ. Χούσος

Εθνικό Κέντρο Τεκμηρίωσης/Εθνικό Ίδρυμα Ερευνών
nhoussos@ekt.gr

Κωνσταντίνος Σταμάτης

Εθνικό Κέντρο Τεκμηρίωσης/Εθνικό Ίδρυμα Ερευνών
kstamatis@ekt.gr

Αλέξανδρος Κουλούρης

Τμήμα Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης, ΤΕΙ Αθήνας
akoul@teiath.gr, koulouris.a@gmail.com

Περίληψη

Με το πέρασμα του χρόνου και την κατακόρυφη αύξηση του όγκου του διαθέσιμου ψηφιακού περιεχομένου, προερχόμενο είτε από πρωτογενώς ηλεκτρονικές πηγές είτε από ψηφιοποίηση, γίνεται ολοένα και μεγαλύτερη η ανάγκη για διαλειτουργικότητα μεταξύ ψηφιακών αποθετηρίων. Η παρούσα εργασία παρουσιάζει επιτυχημένα παραδείγματα διαλειτουργικότητας ελληνικών αποθετηρίων με τη *Europeana* μέσω του έργου *EuropeanaLocal* καθώς και το *DSpace Europeana plug-in*, ένα εργαλείο λογισμικού ανοικτού κώδικα που αναπτύχθηκε από Έλληνες προγραμματιστές και φάνηκε ιδιαίτερα χρήσιμο στην συγκεκριμένη προσπάθεια. Επιπλέον, δίνεται ιδιαίτερη έμφαση στην απαίτηση για επιλεκτική παροχή εγγραφών από αποθετήρια σε συσσωρευτές περιεχομένου και παρουσιάζονται κάποιες επεκτάσεις της πλατφόρμας αποθετηρίων *DSpace* για την παραμετροποίηση και την ευέλικτη διαχείριση των απαραίτητων μηχανισμών και διαδικασιών.

Λέξεις-κλειδιά: Διαλειτουργικότητα, αποθετήρια, *DSpace*, *Europeana*, *OAI-PMH*

1. Εισαγωγή

Η παρούσα εργασία αναφέρεται στη διαλειτουργικότητα αποθετηρίων μέσω του πρωτοκόλλου *OAI-PMH* και σε σχετικά τεχνολογικά εργαλεία, με αναφορά και σε πρόσφατες πραγματικές περιπτώσεις στον ελλαδικό χώρο. Παρουσιάζονται επιτυχημένα παραδείγματα διαλειτουργικότητας ελληνικών αποθετηρίων με τη *Europeana* μέσω του έργου *EuropeanaLocal* καθώς και το *DSpace Europeana plug-in*, ένα εργαλείο λογισμικού ανοικτού κώδικα που αναπτύχθηκε από Έλληνες

προγραμματιστές και φάνηκε ιδιαίτερα χρήσιμο στην συγκεκριμένη προσπάθεια. Επιπλέον, δίνεται ιδιαίτερη έμφαση στην απαίτηση για επιλεκτική παροχή εγγραφών από αποθετήρια σε συσσωρευτές περιεχομένου και παρουσιάζονται κάποιες επεκτάσεις της πλατφόρμας αποθετηρίων DSspace για την παραμετροποίηση και την ευέλικτη διαχείριση των απαραίτητων μηχανισμών και διαδικασιών.

Η διάρθρωση του άρθρου έχει ως εξής: Στην Ενότητα 2 γίνεται μια εισαγωγή σε θέματα διαλειτουργικότητας ψηφιακών αποθετηρίων, αναφέρονται συνοπτικά σημαντικά συστήματα συσσωρευτών από την Ευρώπη και αναπτύσσεται η απαίτηση για επιλεκτική χορήγηση εγγραφών μεταδεδομένων από αποθετήρια προς συσσωρευτές. Στην Ενότητα 3 παρουσιάζεται η μελέτη περίπτωσης ενσωμάτωσης ελληνικών αποθετηρίων στη Europeana μέσω του έργου EuropeanaLocal. Στην Ενότητα 4 παρουσιάζονται διεξοδικά σχετικά τεχνικά ζητήματα και εργαλεία λογισμικού. Η εργασία κλείνει με σύνοψη και συμπεράσματα που περιλαμβάνονται στην Ενότητα 5.

2. Διαλειτουργικότητα σε συστήματα ψηφιακών αποθετηρίων

2.1. Επιθυμητά επίπεδα διαλειτουργικότητας και τεχνικές δυσκολίες

Με το πέρασμα του χρόνου και την κατακόρυφη αύξηση του όγκου του διαθέσιμου ψηφιακού περιεχομένου, προερχόμενο είτε από πρωτογενές ηλεκτρονικές πηγές είτε από ψηφιοποίηση, γίνεται ολοένα και μεγαλύτερη η ανάγκη για διαλειτουργικότητα μεταξύ ψηφιακών αποθετηρίων. Η διαλειτουργικότητα επιτρέπει τη δυνατότητα διάθεσης και γενικότερα αξιοποίησης των μεταδεδομένων επιμέρους αποθετηρίων από τρίτες εφαρμογές, όπως συσσωρευτές περιεχομένου (content aggregators), χωρίς απαίτηση για ad hoc λύσεις ολοκλήρωσης ετερογενών συστημάτων.

Ειδικότερα στα συστήματα αποθετηρίων/ ψηφιακών βιβλιοθηκών, η διαλειτουργικότητα στο πεδίο των μεταδεδομένων μπορεί να θεωρηθεί πως έχει τις παρακάτω τρεις διαστάσεις (Ouksel, 1999):

- **Διαλειτουργικότητα σε επίπεδο συστημάτων αποθετηρίων και ψηφιακών βιβλιοθηκών.** Επιτρέπει την online διάθεση των μεταδεδομένων μιας ψηφιακής βιβλιοθήκης σε τρίτες εφαρμογές και συστήματα. Σημαίνει πρακτικά πως για κάθε επιμέρους σύστημα υπάρχει ένας διαδικτυακός εξυπηρετητής που παρέχει ανά πάσα στιγμή τα μεταδεδομένα του μέσω μιας προγραμματιστικής διεπαφής που υλοποιείται βάσει ενός συγκεκριμένου πρωτοκόλλου διάθεσης. Συνήθως η online διάθεση μπορεί να πραγματοποιηθεί είτε σε επίπεδο πλήρους διαμοιρασμού ή συγκομιδής (harvesting), όπου πρακτικά μεταφορτώνεται το σύνολο (ή προκαθορισμένα υποσύνολα) των εγγραφών ενός συστήματος για

επαναχρησιμοποίηση σε άλλες εφαρμογές, είτε σε επίπεδο μετα-αναζήτησης (meta-search) όπου ανακτώνται μόνο εγγραφές που ταιριάζουν με συγκεκριμένα κριτήρια αναζήτησης. Η διαλειτουργικότητα σε επίπεδο συστημάτων περιλαμβάνει και την κατάλληλη κωδικοποίηση (π.χ. Unicode) των μεταδεδομένων που ανταλλάσσονται.

- **Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής.** Επιτρέπει σε διαφορετικά συστήματα να «διαβάζουν» σωστά τα δεδομένα που ανταλλάσσουν μεταξύ τους. Στην ιδανική περίπτωση, αυτό προϋποθέτει τα εξής:
 - Χρήση κοινής γλώσσας για την κωδικοποίηση των μεταδεδομένων σε συντακτικό επίπεδο (π.χ. XML).
 - Κωδικοποίηση των τιμών των δεδομένων σύμφωνα με ένα κοινό πρότυπο. Για παράδειγμα, οι ημερομηνίες να αναγράφονται με τον ίδιο τρόπο σε όλα τα συστήματα.
 - Χρήση κοινών σχημάτων μεταδεδομένων (π.χ. Dublin Core, MODS, CDWA, EAD, κλπ.)
- **Διαλειτουργικότητα σε σημασιολογικό επίπεδο.** Επιτρέπει σε διαφορετικά συστήματα να «κατανοούν» σωστά τη σημασία των δεδομένων που ανταλλάσσουν μεταξύ τους. Για την πλήρη διαλειτουργικότητα σε σημασιολογικό επίπεδο, κάθε πεδίο μεταδεδομένων που ανταλλάσσεται χρειάζεται να έχει δηλωμένη, σαφή σημασία. Για παράδειγμα, όταν για ένα έργο τέχνης έχουμε ένα πεδίο date, τι ακριβώς εννοούμε με αυτό (π.χ. πρόκειται για την ημερομηνία δημιουργίας, πρώτης έκθεσης στο κοινό, πρόσκτησης από το μουσείο που το φιλοξενεί ή εισαγωγής στην ψηφιακή βιβλιοθήκη;). Η χρήση συμβατών όρων σε ελεγχόμενα λεξιλόγια και θησαυρούς είναι μία ακόμα σημαντική πτυχή της σημασιολογικής διαλειτουργικότητας. Το ιδανικό είναι η δήλωση της σημασίας κάθε στοιχείου να πραγματοποιείται μέσω μιας κατάλληλης γλώσσας αναπαράστασης γνώσης όπως RDF ή OWL.

Στη διεθνή κοινότητα των ψηφιακών βιβλιοθηκών και γενικότερα της διαχείρισης πληροφορίας υπάρχει έντονη δραστηριότητα τουλάχιστον από τις αρχές της δεκαετίας του 1990 σχετικά με την επίτευξη διαλειτουργικότητας στα παραπάνω επίπεδα. Σήμερα μπορεί κανείς να θεωρήσει ότι η συγκεκριμένη προσπάθεια έχει εξαιρετικά αποτελέσματα με πλήρη πρακτική εφαρμογή σε πραγματικά συστήματα μόνο όσον αφορά τη διαλειτουργικότητα συστημάτων και κάποιες πλευρές της διαλειτουργικότητας σε επίπεδο σύνταξης και δομής. Για παράδειγμα, μεγάλο ποσοστό των σύγχρονων αποθετηρίων υποστηρίζει το πρωτόκολλο συγκομιδής μεταδεδομένων OAI-PMH, ενώ τα μεταδεδομένα είναι

κατά κανόνα διαθέσιμα σε μορφή XML και υποστηρίζονται τουλάχιστον κάποια πεδία του απλού Dublin Core. Παρ' όλα αυτά, η σημασιολογική διαλειτουργικότητα σε μεγάλης κλίμακας συστήματα φαίνεται να είναι αρκετά μακριά χρονικά, ενώ υπάρχουν ακόμα προβλήματα σε ζητήματα όπως η κωδικοποίηση των τιμών των δεδομένων και η ευρεία χρήση κοινών σχημάτων μεταδεδομένων πιο πολύπλοκων και με μεγαλύτερη εκφραστική ικανότητα και σαφήνεια από το Dublin Core.

2.2. Το διεθνές σκηνικό σε συστήματα συσσωρευτών

Σύμφωνα με τη διεθνή εμπειρία, η ύπαρξη κεντρικών συστημάτων συσσώρευσης ψηφιακού περιεχομένου (content aggregator) σε επίπεδο κρατικό ή και διεθνές αυξάνει σημαντικά την αποδοτικότητα των πόρων που επενδύονται σε έργα ψηφιοποίησης και παραγωγής περιεχομένου, με τη διασφάλιση της ευρύτερης δυνατής πρόσβασης στα αποτελέσματά τους, κυρίως όμως, με την προσφορά νέων υπηρεσιών και μοντέλων επαναχρησιμοποίησης του περιεχομένου. Στους συσσωρευτές συγκεντρώνεται με τη μέθοδο της συγκομιδής (harvesting) – με περιοδική αυτόματη ανανέωση σε τακτά χρονικά διαστήματα – το σύνολο του εγγραφών μεταδεδομένων των επιμέρους συστημάτων. Συνήθως οι συσσωρευτές δεν συλλέγουν το ίδιο ψηφιακό υλικό (παρά μόνο ίσως κάποια αναπαράσταση χρήσιμη για συνοπτική παρουσίαση, π.χ. ένα thumbnail στην περίπτωση εικόνων) και, όσον αφορά τα μεταδεδομένα, ανακτούν ένα μέρος της εγγραφής. Για να δει την πλήρη εγγραφή και το ίδιο το τεκμήριο στη υψηλότερη δημόσια διαθέσιμη ανάλυση, ο χρήστης οδηγείται από τη δικτυακή πύλη του συσσωρευτή στην αντίστοιχη σελίδα του αποθετηρίου.

Ειδικότερα, οι συσσωρευτές παρέχουν συνήθως τις εξής λειτουργίες:

- Περιοδική συγκομιδή μεταδεδομένων από τα επιμέρους αποθετήρια.
- Διαδικτυακή πύλη μέσω της οποίας παρέχονται δυνατότητες ενιαίας αναζήτησης και πλοήγησης στο σύνολο του περιεχομένου με διάφορα κριτήρια.
- Μοναδικό κεντρικό σημείο για τη διάθεση του συνόλου του συλλεχθέντος περιεχομένου μέσω web services σε τρίτα συστήματα για αξιοποίηση σε ανάπτυξη νέων υπηρεσιών / εφαρμογών.
- Αυτόματο έλεγχο για τη συμβατότητα των εγγραφών της κάθε επιμέρους ψηφιακής βιβλιοθήκης / αποθετηρίου με τις προδιαγραφές – πιθανόν και με αναλυτική αναφορά των τυχόν ελλείψεων.

- Παραγωγή συγκεντρωτικών στοιχείων αναφορών για το σύνολο του ψηφιακού περιεχομένου και την πρόσβαση σε αυτό από χρήστες. Προς το παρόν αυτή η λειτουργία αφορά στατιστικά στοιχεία πρόσβασης χρηστών στις εγγραφές μόνο μέσω της διαδικτυακής πύλης του συσσωρευτή. Επιπλέον, υπάρχουν προσπάθειες διεθνώς για την ανάπτυξη μηχανισμών κεντρικής συγκέντρωσης στατιστικών στοιχείων πρόσβασης σε περιεχόμενο από τα επιμέρους αποθετήρια και προβολής τους από τη διαδικτυακή πύλη του συσσωρευτή.

Σήμερα υπάρχει πληθώρα συσσωρευτών που λειτουργούν διεθνώς. Σε εθνικό επίπεδο στην Ελλάδα είναι ιδιαίτερα αξιόλογη η προσπάθεια που αφορά το openarchives.gr, συσσωρευτή ελληνικών ψηφιακών βιβλιοθηκών και αποθετηρίων που έχει υιοθετήσει την προσέγγιση της κεντρικής συγκομιδής μεταδεδομένων (βασικών πεδίων του Dublin Core) μέσω OAI-PMH και αυτή τη στιγμή αποτελεί πύλη για 42 ελληνικά συστήματα και 361323 εγγραφές και παρέχει λειτουργίες ενιαίας αναζήτησης και πλοήγησης στο συγκεκριμένο περιεχόμενο.

Παραδείγματα σημαντικών συστημάτων υπερ-εθνικής εμβέλειας από τον ευρωπαϊκό χώρο είναι τα παρακάτω:

Europeana. Η Europeana είναι μια πρωτοβουλία της Ευρωπαϊκής Επιτροπής, με στόχο να αναπτύξει ένα κοινό σημείο πρόσβασης για την ψηφιακή πολιτιστική και επιστημονική κληρονομιά της Ευρώπης. Η Europeana φιλοξενεί σήμερα βιβλία, χάρτες, ηχογραφήσεις, φωτογραφίες, αρχαιακά έγγραφα, πίνακες και ταινίες, από βιβλιοθήκες, αρχεία και πολιτιστικά ιδρύματα των 27 κρατών μελών της Ευρωπαϊκής Ένωσης. Το περιεχόμενο συγκεντρώνεται από συσσωρευτές και παρόχους περιεχομένου από όλη την Ευρώπη. Το πρωτόκολλο συλλογής είναι το OAI-PMH και το σχήμα μεταδεδομένων το Europeana Semantic Elements (ESE) που είναι βασισμένο στο Dublin Core. Για την εισαγωγή περιεχομένου υπάρχουν εργαλεία διαθέσιμα στο εξουσιοδοτημένο προσωπικό της Europeana για αυτοματοποιημένο έλεγχο του προς εισαγωγή περιεχομένου (Europeana content checker). Περισσότερες πληροφορίες για τη Europeana και για την επιτυχημένη ενσωμάτωση σε αυτή συλλογών ελληνικών φορέων μέσω του έργου EuropeanaLocal – με συντονιστή από την Ελλάδα τη Δημόσια Κεντρική Βιβλιοθήκη της Βέροιας – περιλαμβάνονται στην Ενότητα 3 του παρόντος άρθρου.

Δύο ακόμη πολύ σημαντικά έργα χρηματοδοτούμενα από την Ευρωπαϊκή Ένωση με κύριο αντικείμενο την ενσωμάτωση περιεχομένου στη Europeana, είναι το ATHENA και το CARARE.

Το έργο ATHENA αποσκοπεί στη συνεργασία μουσείων και οργανισμών πολιτιστικής κληρονομιάς για την ενσωμάτωση του υλικού τους στη Europeana καθώς και την ανάπτυξη σχετικών εργαλείων που θα διευκολύνουν στο μέλλον την τροφοδότηση της συγκεκριμένης ευρωπαϊκής ψηφιακής βιβλιοθήκης με παρόμοιο υλικό από αντίστοιχους φορείς. Στο έργο συμμετέχουν από την Ελλάδα το Εθνικό Μετσόβιο Πολυτεχνείο, το Υπουργείο Πολιτισμού και Τουρισμού και το Πανεπιστήμιο Πατρών. Η προσέγγιση που ακολουθείται είναι η συγκέντρωση του περιεχομένου όλων των παρόχων σε ενδιάμεσο σύστημα το οποίο δεν λειτουργεί ως αυτόνομος συσσωρευτής (δεν είναι ανοικτός για harvesting από τρίτους, παρά μόνο από τη Europeana) και παρέχει εργαλεία για τη φόρτωση μεταδεδομένων σε μορφή XML και το μετασχηματισμό τους (mapping) στο σχήμα μεταδεδομένων που υποστηρίζεται από τη Europeana. Σημειώνεται ότι το συγκεκριμένο σύστημα και τα αντίστοιχα εργαλεία έχουν αναπτυχθεί στο πλαίσιο του έργου από το Εργαστήριο Ψηφιακής Επεξεργασίας Σημάτων και Εικόνων του Εθνικού Μετσόβιου Πολυτεχνείου.

Το CARARE έχει ως στόχο τον εμπλουτισμό της Europeana με περιεχόμενο που αφορά αρχαιολογικά μνημεία και την αρχιτεκτονική κληρονομιά της Ευρώπης, καθώς και την προσπέλαση σε υλικό 3D και εικονικής πραγματικότητας μέσα από τη Europeana. Στο έργο, το οποίο έχει ξεκινήσει πρόσφατα (Φεβρουάριος 2010), συμμετέχουν από ελληνικής πλευράς το ερευνητικό κέντρο Αθηνά, το Εθνικό Μετσόβιο Πολυτεχνείο και το Υπουργείο Πολιτισμού και Τουρισμού.

DRIVER. Πρόκειται για μεγάλης κλίμακας υποδομή για τη διαλειτουργικότητα αποθετηρίων, η οποία με πυρήνα έναν κεντρικό συσσωρευτή επιτρέπει πρόσβαση σε όλες τις εγγραφές μέσω μιας κεντρικής διαδικτυακής πύλης. Το πρωτόκολλο συλλογής είναι το OAI-PMH, ενώ το σχήμα μεταδεδομένων είναι βασισμένο στο Dublin Core. Υπάρχουν αναλυτικές προδιαγραφές για τη συμβατότητα αποθετηρίων με την κεντρική υποδομή, η οποία περιλαμβάνει επίσης εργαλεία για διαχειριστές αποθετηρίων για διεξαγωγή αυτοματοποιημένων ελέγχων συμβατότητας με τις συγκεκριμένες προδιαγραφές.

OpenAIRE. Το OpenAIRE αναπτύσσει τις κατάλληλες τεχνολογικές υποδομές και τους μηχανισμούς υποστήριξης για τον εντοπισμό, την απόθεση, την πρόσβαση και την παρακολούθηση επιστημονικών δημοσιεύσεων που έχουν προέλθει από ερευνητικά έργα χρηματοδοτούμενα από την Ευρωπαϊκή Ένωση και ειδικότερα συγκεκριμένες θεματικές περιοχές του FP7 και τα προγράμματα του Ευρωπαϊκού Συμβουλίου Έρευνας – ERC. Οι δημοσιεύσεις θα είναι προσβάσιμες μέσω κεντρικής δικτυακής πύλης, ενώ για τη συλλογή των μεταδεδομένων θα υιοθετηθεί προσέγγιση της ίδιας φιλοσοφίας με το DRIVER.

Συντονιστής του έργου είναι το Πανεπιστήμιο Αθηνών, Τμήμα Πληροφορικής και Τηλεπικοινωνιών, ενώ από την Ελλάδα συμμετέχει και το Εθνικό Κέντρο Τεκμηρίωσης.

DART Europe. Πρόκειται για κοινοπραξία που διατηρεί την πανευρωπαϊκή δικτυακή πύλη / συσσωρευτή διδακτορικών διατριβών (www.dart-europe.eu). Η συλλογή πραγματοποιείται μέσω ΟΑΙ-ΡΜΗ και απαιτούνται κάποια βασικά στοιχεία ανά εγγραφή διατριβής σε μορφή Dublin Core. Από ελληνικής πλευράς συμμετέχει ως εταίρος το ΕΚΤ που παρέχει εγγραφές από το αποθετήριο του Εθνικού Αρχείου Διδακτορικών Διατριβών.

2.3. Το πρόβλημα για τους διαχειριστές αποθετηρίων

Από την παράγραφο 2.2 γίνεται κατανοητό ότι σήμερα οι διαχειριστές αποθετηρίων έχουν να αντιμετωπίσουν το ζήτημα της συμβατότητας των συστημάτων για τα οποία είναι υπεύθυνοι με εθνικές και διεθνείς υποδομές συσσωρευτών όχι ως ένα δευτερεύον θέμα θεωρούμενο ως πολυτέλεια, αλλά ως μια στοιχειώδη απαίτηση για κάθε υποδομή αποθετηρίου (π.χ. αποθετήρια ιδρυμάτων, θεματικά αποθετήρια, ψηφιακές συλλογές πολιτιστικού περιεχομένου).

Τα κύρια θέματα που έχουν να αντιμετωπίσουν οι διαχειριστές αποθετηρίων όσον αφορά τη διαλειτουργικότητα των συστημάτων τους με τους συσσωρευτές είναι τα εξής:

1. Υποστήριξη του ΟΑΙ-ΡΜΗ ως πρωτοκόλλου συγκομιδής.
2. Συμβατότητα του σχήματος μεταδεδομένων του αποθετηρίου με αυτό που προδιαγράφει ο εκάστοτε συσσωρευτής και υλοποίηση των απαραίτητων μετατροπών.
3. Συμβατότητα των κωδικοποιήσεων τιμών των δεδομένων με αυτές που υποστηρίζει ο εκάστοτε συσσωρευτής και υλοποίηση των απαραίτητων μετατροπών, αν είναι εφικτό.
4. Επιλεκτική παροχή στους συσσωρευτές εγγραφών του αποθετηρίου.

Στο παρόν θα μας απασχολήσει ιδιαίτερα από τον παραπάνω κατάλογο το σημείο 4, η δυνατότητα δηλαδή ενός αποθετηρίου να παρέχει για συγκομιδή σε ένα συσσωρευτή όχι το σύνολο των εγγραφών του αλλά επιλεγμένα υποσύνολα με βάση διάφορα κριτήρια. Η δυνατότητα αυτή είναι απαραίτητη στην πράξη, καθώς σε πολλές περιπτώσεις ένας συσσωρευτής δεν επιθυμεί να συλλέξει όλες τις εγγραφές ενός αποθετηρίου. Κάτι τέτοιο μπορεί να συμβαίνει για διάφορους λόγους, ενδεικτικά αναφέρονται οι παρακάτω:

- Ο συσσωρευτής συλλέγει αποκλειστικά εγγραφές που πληρούν συγκεκριμένα κριτήρια ως προς το θέμα των δυνατοτήτων ανοικτής πρόσβασης και των πνευματικών δικαιωμάτων, όπως:
 - Μόνο εγγραφές για τις οποίες υπάρχει ψηφιακό τεκμήριο (π.χ. το πλήρες κείμενο για επιστημονικά άρθρα, βιβλία, κλπ.) και είναι δημόσια διαθέσιμο μέσω του επιμέρους αποθετηρίου. Παρόμοια πολιτική ακολουθούν το DRIVER, το OpenAIRE και το DART Europe.
 - Μόνο εγγραφές τα μεταδεδομένα των οποίων διατίθενται ελεύθερα για διάφορες χρήσεις, ιδανικά μέσω κατάλληλης άδειας, π.χ. κάποια από τις άδειες της οικογένειας Creative Commons. Προδιαγραφές αυτού του τύπου θέτει, για παράδειγμα, η Europeana.
- Ο συσσωρευτής συλλέγει αποκλειστικά εγγραφές για τεκμήρια συγκεκριμένων θεματικών περιοχών, ενώ τα επιμέρους αποθετήρια μπορεί να είναι διεπιστημονικά. Τέτοια περίπτωση είναι το αποθετήριο HealthCompetence.eu με αποτελέσματα ερευνών χρηματοδοτούμενων από την Ευρωπαϊκή Ένωση στο χώρο της υγείας.
- Ο συσσωρευτής συλλέγει αποκλειστικά εγγραφές για τεκμήρια συγκεκριμένου τύπου (π.χ. διδακτορικές διατριβές, όπως το DART Europe), ενώ τα επιμέρους αποθετήρια μπορεί να περιέχουν υλικό διάφορων τύπων.

Από τα παραπάνω καταδεικνύεται ότι η πολυπλοκότητα της επιλεκτικής παροχή εγγραφών σε συσσωρευτές. Η απαίτηση αυτή γίνεται μάλιστα πιο δύσκολο να υλοποιηθεί αν αναλογιστεί κανείς πως ένα αποθετήριο είναι πολύ πιθανό να παρέχει εγγραφές σε περισσότερους από έναν συσσωρευτές, με διαφορετικές προδιαγραφές για το καθένα. Σημειώνεται πως ο μηχανισμός των προκαθορισμένων συνόλων (sets) του OAI (βλ. Ενότητα 4) που συνήθως υλοποιείται σε πλατφόρμες αποθετηρίων σε αντιστοιχία με επιμέρους κοινότητες/συλλογές του αποθετηρίου (π.χ. εξ ορισμού ένα σύνολο ανά συλλογή) δεν επαρκεί καθώς, όπως είναι προφανές από τα προαναφερθέντα παραδείγματα, τα επιθυμητά σύνολα προς συγκομιδή μπορεί να περιέχουν εγγραφές από διαφορετικές συλλογές. Η επιλεκτική ανάκτηση εγγραφών με προσδιορισμό πολύπλοκων κριτηρίων από την πλευρά του συσσωρευτή (όχι προκαθορισμένων από τα επιμέρους αποθετήρια) είναι μια δυνατότητα που θα διευκόλυνε πολύ το έργο των συσσωρευτών αλλά και των διαχειριστών αποθετηρίων, αυτή τη στιγμή όμως λείπει από το OAI-PMH (Mazurek, 2009).

3. Μελέτη περίπτωσης: EuropeanaLocal

3.1. Εισαγωγή

Μέχρι τον Ιούνιο 2010, επτά ελληνικοί πολιτιστικοί φορείς αποτελούν την πρώτη ομάδα οργανισμών που έχουν ήδη εντάξει και αναδεικνύουν το ψηφιακό περιεχόμενό τους μέσω της Ψηφιακής Πύλης Europeana (Koninklijke Bibliotheek, 2009):

- Αμερικάνικη Γεωργική Σχολή - Βιβλιοθήκη «Δημήτρης & Αλίκη Περρωτή»
- Δημόσια Κεντρική Βιβλιοθήκη Βέροιας
- Εθνικό Κέντρο Τεκμηρίωσης
- Εθνικό Μετσόβιο Πολυτεχνείο
- Μέγαρο Μουσικής – Μεγάλη Μουσική Βιβλιοθήκη της Ελλάδος «Λίλιαν Βουδούρη»
- Τεχνικό Επιμελητήριο Ελλάδας – Περιφερειακό Τμήμα Κέρκυρας Κέρκυρας
- openarchives.gr (ως μηχανή αναζήτησης ελληνικών ψηφιακών βιβλιοθηκών)

Η Δημόσια Κεντρική Βιβλιοθήκη Βέροιας (ΔΚΒΒ), είναι ο εθνικός συσσωρευτής (aggregator) μεταδεδομένων, τουλάχιστον μέχρι τη λήξη του προγράμματος (Ιούνιο 2011) και το openarchives.gr, είναι μια μηχανή αναζήτησης ελληνικών ψηφιακών πηγών.

Οι ελληνικοί φορείς παροχής περιεχομένου (content providers) και ο συσσωρευτής μεταδεδομένων τους, η ΔΚΒΒ, έχουν επιτυχώς διεξάγει με επιτυχία μια δοκιμή αυτόματης συγκομιδής μεταδεδομένων. Στο κεφάλαιο αυτό αναλύονται οι πρακτικές που έχουν εφαρμόσει τα ελληνικά ψηφιακά αποθετήρια, προκειμένου να είναι συμβατά με το προφίλ μεταδεδομένων Dublin Core (DC) που έχει επιλέξει η Europeana, το *Europeana Semantic Elements* (ESE) (2009), έτσι ώστε στη συνέχεια να εξάγουν τα ESE μεταδεδομένα τους, με τη χρήση του OAI-PMH, στη ΔΚΒΒ, όπως ήδη έχει γίνει. Σ' αυτή τη φάση, ισχύουν οι ακόλουθες συμφωνίες παροχής μεταδεδομένων. Οι συμφωνίες είναι δύο ειδών και αφορούν τους παρόχους περιεχομένου, αλλά και τους εθνικούς φορείς συγκέντρωσης μεταδεδομένων (national aggregators). Το σχήμα που έχει εφαρμοστεί με επιτυχία είναι το ακόλουθο. Οι πάροχοι θα υπογράψουν διμερείς συμφωνίες με τον εθνικό φορέα συσσωρευτή μεταδεδομένων, δηλαδή την ΔΚΒΒ, τουλάχιστον για τη διάρκεια του EuropeanaLocal (Ιούνιος 2008 – Ιούνιος 2011). Επομένως, οι ελληνικοί

φορείς περιεχομένου υπογράφουν συμφωνίες με τη ΔΚΒΒ και έχουν την υποχρέωση να εξάγουν τα ESE μεταδεδομένα τους στη ΔΚΒΒ μέσω ΟΑΙ-ΡΜΗ. Η ΔΚΒΒ έχει την υποχρέωση να συγκεντρώνει τα ESE μεταδεδομένα των φορέων, δηλαδή να κάνει την αυτόματη συγκομιδή τους (harvesting). Στη συνέχεια η ΔΚΒΒ υπογράφει άλλη συμφωνία με την Europeana και έχει την υποχρέωση να παρέχει στη Europeana τα μεταδεδομένα που έχει συγκεντρώσει με «harvesting». Τέλος, η Europeana με «harvesting» θα συλλέγει τα μεταδεδομένα από όλους τους εθνικούς «aggregators» και από τη Βέροια. Η επιτυχημένη δοκιμή έγινε στα μέσα Απριλίου 2010 και ήδη, από τον Ιούνιο 2010, τα μεταδεδομένα των φορέων είναι διαθέσιμα από την Europeana.

3.2. Εναλλακτικά σενάρια που υλοποιήθηκαν μέχρι σήμερα για την αυτόματη συγκομιδή μεταδεδομένων

Στο πρόγραμμα EuropeanaLocal, μπορούν να συμμετέχουν βιβλιοθήκες, αρχεία, μουσεία και άλλοι φορείς πολιτισμού με ψηφιοποιημένο περιεχόμενο (κείμενο, φωτογραφίες, βίντεο, ήχο, κ.λπ.), που να έχουν διάθεση για συνεργασία και να τηρούν κάποια πρότυπα διαλειτουργικότητας (π.χ. το πρωτόκολλο ΟΑΙ-ΡΜΗ) (Lagoze et al 2002). Παρουσιάζονται τρία εναλλακτικά σενάρια για τη συμμετοχή των ελληνικών φορέων πολιτισμού στη Europeana, αξιοποιώντας την υποδομή και την τεχνογνωσία που αναπτύσσεται στο πλαίσιο του EuropeanaLocal (2008).

Το πρώτο σενάριο αφορά στην περίπτωση που ο φορέας έχει αποθετήριο και ψηφιακό περιεχόμενο. Τότε, αρκεί να ακολουθήσει το ESE, έτσι ώστε να είναι δυνατή η αυτόματη συγκομιδή μεταδεδομένων και να ενεργοποιηθεί το πρωτόκολλο ΟΑΙ-ΡΜΗ, κάτι το οποίο είναι εύκολο στην περίπτωση που οι φορείς έχουν χρησιμοποιήσει λογισμικό ανοικτού κώδικα (π.χ. DSpace, FEDORA, OJS) για το αποθετήριό τους. Ειδικά για την δημιουργία των ESE αλλά και για την εξαγωγή τους μέσω ΟΑΙ-ΡΜΗ στον εθνικό φορέα συγκέντρωσης μεταδεδομένων, που στα πλαίσια του EuropeanaLocal είναι η ΔΚΒΒ, η ελληνική ομάδα διαχείρισης του έργου (Δημόσια Κεντρική Βιβλιοθήκη Βέροιας, 2009), έχει αναπτύξει τα κατάλληλα εργαλεία, τόσο σε επίπεδο συσσωρευτή μεταδεδομένων, όσο και επίπεδο μετατροπής σε ESE και εξαγωγής τους με ΟΑΙ-ΡΜΗ στη ΔΚΒΒ.

Ειδικά δε για όσους φορείς έχουν DSpace, η ομάδα έχει αναπτύξει ένα εργαλείο σε PHP και σε Java (Banos, 2010), το οποίο ενημερώνει αυτόματα όχι μόνο το σχήμα μεταδεδομένων, αλλά και τις τιμές σε κάθε πεδίο του σχήματος. Επιπλέον, έχει επιλυθεί και το θέμα της μετατροπής του ΟΑΙ-ΡΜΗ σε

εγκαταστάσεις DSpace, έτσι ώστε να εξάγει όχι μόνο DC μεταδεδομένα, αλλά και ESE. Σχετική υλοποίηση έχει γίνει στη Μέδουσα, στο αποθετήριο της Δημόσιας Κεντρικής Βιβλιοθήκης Βέροιας (2010), βλ. <http://medusa.libver.gr/oai/request?verb=ListRecords&metadataPrefix=ese> για τα ESE και

http://medusa.libver.gr/oai/request?verb=ListRecords&metadataPrefix=oai_dc για τα DC μεταδεδομένα.

Τέλος, η ΔΚΒΒ έχει αναπτύξει το συσσωρευτή (Veria Central Public Library, 2010), ο οποίος ενσωματώνει με αυτόματο τρόπο ESE μεταδεδομένα από τους ελληνικούς φορείς που συμμετέχουν στη Europeana, και αφού κάνει κάποια τεστ διπλοεγγραφών τα εξάγει στη Europeana. Όλοι οι φορείς που θέλουν να συμμετέχουν στο EuropeanaLocal, πρέπει να ελεγχθούν από τον συσσωρευτή και να κάνουν τεστ αυτόματης συγκόμιδης.

Αν λοιπόν πληρούνται αυτές οι ελάχιστες προϋποθέσεις, τότε η συμμετοχή είναι δεδομένη, αρκεί να υπάρχει διάθεση και καλή θέληση. Επιπλέον, δεσμεύσεις οικονομικής και νομικής φύσεως ή υποχρεώσεις δεν υπάρχουν. Ειδικά δε στην Europeana δεν τίθεται θέμα προβλημάτων που απορρέουν από τα δικαιώματα πνευματικής ιδιοκτησίας του περιεχομένου, διότι μόνο τα μεταδεδομένα συλλέγονται με αυτόματο τρόπο και γι' αυτά αναπτύσσονται συμφωνίες που περιγράφηκαν παραπάνω. Το περιεχόμενο δεν μεταναστεύει, αλλά παραμένει στο κατανομημένο αποθετήριο και ο φορέας, αποφασίζει την πολιτική διάθεσής του, ανάλογα με τις ανάγκες του.

Το δεύτερο σενάριο αφορά στην περίπτωση που ο φορέας δεν έχει αποθετήριο και ψηφιακό περιεχόμενο. Στο σενάριο αυτό υπάρχουν δύο διαθέσιμες επιλογές. Η πρώτη και πιο χρονοβόρα, είναι η από το μηδέν σύσταση αποθετηρίου με τη χρήση των προτύπων που προτείνει η Europeana. Αυτό μπορεί να απαιτεί επιπλέον εργασία, χρόνο και κόπο, αλλά έχει το πλεονέκτημα ότι και το περιεχόμενο και το αποθετήριο θα χτιστεί με πρότυπα και θα είναι διαλειτουργικό χωρίς να χρειάζεται επιπλέον παραμετροποίηση. Η δεύτερη επιλογή, είναι η προσωρινή διάθεση του περιεχομένου του φορέα, μέσω του αποθετηρίου της ΔΚΒΒ. Αυτό μέχρι ο εκάστοτε φορέας να υλοποιήσει το δικό του αποθετήριο και με την προϋπόθεση ότι η συλλογή του φιλοξενείται στο αποθετήριο της ΔΚΒΒ και φαίνεται στη Europeana, ως συλλογή του φορέα και όχι ως συλλογή της ΔΚΒΒ. Πρέπει να γίνει κατανοητό ότι υπάρχει αρκετός χρόνος για να διαμορφωθούν οι συνθήκες ωρίμανσης για τη συμμετοχή των ελληνικών πολιτιστικών οργανισμών στο πρόγραμμα EuropeanaLocal μιας και αυτό λήγει το 2011.

Το τρίτο και τελευταίο σενάριο αφορά στην περίπτωση που ο φορέας έχει κάποιο περιεχόμενο (π.χ. συλλογή εικόνων) στον δικτυακό του τόπο, το οποίο μπορεί να εκφραστεί σε XML. Επίσης, ακόμα και στην περίπτωση που κάποιος φορέας δεν χρησιμοποιεί λογισμικό ανοικτού κώδικα ή λογισμικό που δεν υποστηρίζει OAI-PMH ή οποιαδήποτε άλλη μορφή αυτόματης εξαγωγής και ανταλλαγής μεταδεδομένων, υπάρχει το εργαλείο ανοικτού κώδικα ΔΕΙΧΤο (Donas, 2010), το οποίο στηρίζεται στο W3C Document Object Model (DOM) και επιτρέπει την εξαγωγή από αδόμητο περιεχόμενο διαδικτυακών τόπων δομημένων μεταδεδομένων σε ESE και σε DC που μπορούν να συλλεχθούν από οποιοδήποτε συσσωρευτή. Επομένως, ακόμα και όταν δεν υπάρχει περιεχόμενο σε πρότυπη μορφή (Dublin Core), διαθέσιμο μέσω του πρωτόκολλου OAI-PMH, είναι εφικτή η συμμετοχή.

Γίνεται κατανοητό από τα παραπάνω ότι οι προϋποθέσεις για τη συμμετοχή των ελληνικών πολιτιστικών οργανισμών σε ένα τέτοιο εγχείρημα είναι λίγες και εύκολο να υλοποιηθούν. Το όφελος είναι μεγάλο: προβολή, συνεργασία, προτυποποίηση, διαλειτουργικότητα, τεχνογνωσία, κ.λπ. Η πρόκληση είναι εδώ. Το ερώτημα είναι αν υπάρχει η βούληση για συμμετοχή σε ένα τέτοιο εγχείρημα.

4. Εργαλεία και μηχανισμοί για παροχή μεταδεδομένων με OAI-PMH

Στο κεφάλαιο αυτό, θα παρουσιάσουμε δύο εργαλεία / μηχανισμούς για την συγκομιδή μεταδεδομένων από αποθετήρια συμβατά με OAI-PMH που χρησιμοποιούν το λογισμικό ανοικτού κώδικα DSpace. Οι δύο μηχανισμοί είναι γενικής χρήσης, αλλά έχουν εφαρμοστεί στις συγκεκριμένες περιπτώσεις για παροχή μεταδεδομένων στη Europeana. Πριν την αναλυτική παρουσίαση των μηχανισμών (παράγραφοι 4.3 και 4.4) γίνεται μια σύντομη εισαγωγή στο πρωτόκολλο OAI-PMH (4.1) και στην υποστήριξή του στο DSpace (4.2).

4.1. Εισαγωγή

Το πρωτόκολλο OAI-PMH αποτελεί μια ανεξάρτητη πλατφόρμα διαλειτουργικότητας για την συγκομιδή μεταδεδομένων από αποθετήρια ανοικτής πρόσβασης, όπου υπάρχει επικοινωνία μεταξύ δύο οντοτήτων: Ο συλλέκτης (harvester) είναι μια εφαρμογή πελάτη η οποία θέτει OAI-PMH αιτήματα στον πάροχο δεδομένων ενός αποθετηρίου προκειμένου να συλλέξει μεταδεδομένα. Από την άλλη πλευρά, ένα αποθετήριο δύναται να

επεξεργαστεί τα ΟΑΙ-ΡΜΗ αιτήματα και να παρέχει μεταδεδομένα μέσω του λεγόμενου πάροχου δεδομένων.

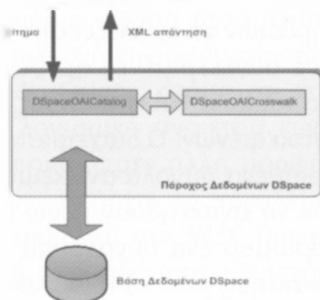
Η επιλεκτική συγκομιδή δεδομένων από το ΟΑΙ-ΡΜΗ μπορεί να γίνει με 2 τρόπους (Lagoze et al 2002). Ο ένας βασίζεται σε χρονικούς δείκτες και επιτρέπει στον πάροχο υπηρεσιών να συλλέξει από το αποθετήριο αντικείμενα που έχουν δημιουργηθεί σε συγκεκριμένες ημερομηνίες. Ο άλλος τρόπος, ο οποίος μας ενδιαφέρει στο πλαίσιο του παρόντος άρθρου, είναι η ανάκτηση μόνο ενός συνόλου (set) αντικειμένων. Ο διαχειριστής του αποθετηρίου μπορεί να δημιουργήσει προκαθορισμένα σύνολα αντικειμένων που διακρίνονται από το όνομά τους και μπορούν να ανακτηθούν μέσω ΟΑΙ-ΡΜΗ βάσει αυτού. Στο DSpace δημιουργείται εξ ορισμού ένα σύνολο για κάθε κοινότητα / συλλογή του αποθετηρίου, άρα η ανάκτηση από κάποιο συσσωρευτή αποκλειστικά των περιεχομένων μίας κοινότητας / συλλογής είναι πολύ εύκολη, κάτι που δεν ισχύει για την επιλεκτική ανάκτηση αντικειμένων από περισσότερες από μία κοινότητες / συλλογές.

4.2. DSpace πάροχος δεδομένων

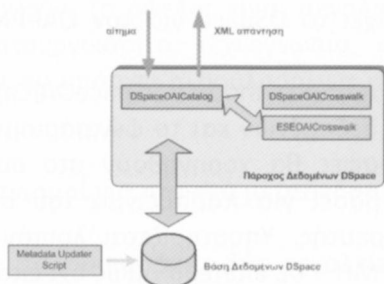
Η υλοποίηση που παρέχει το DSpace για τον ΟΑΙ-ΡΜΗ πάροχο δεδομένων φαίνεται στην **Εικόνα 1**.

Τα αιτήματα λαμβάνονται από το module DSpaceOAIcatalog όπου και γίνεται η ανάκτηση από τη βάση δεδομένων και το φιλτράρισμα των εγγραφών – εδώ καθορίζεται ποιες εγγραφές θα χορηγηθούν στο συσσωρευτή στο τρέχον αίτημα και ποιες όχι, βάσει για παράδειγμα του συνόλου (set) που έχει προσδιορίσει ο συσσωρευτής. Υποστηρίζεται λοιπόν επιλεκτική συγκομιδή όπως ορίζεται στο ΟΑΙ-ΡΜΗ – σε επίπεδο όμως όχι επαρκές για τις απαιτήσεις των συσσωρευτών, όπως έχει αναλυθεί στο Κεφάλαιο 2.

Στη συνέχεια, και πρώτου τα δεδομένα εξαχθούν, ένα δεύτερο module με όνομα DSpaceOAIclosswalk αναλαμβάνει να φιλτράρει τις εγγραφές σε επίπεδο πεδίων (εδώ καθορίζεται ποια πεδία θα χορηγηθούν ανά εγγραφή και ποια όχι) και να παράγει εγγραφές σύμφωνες με το πρότυπο ΟΑΙ-ΡΜΗ.



Εικόνα 1. Πάροχος δεδομένων DSpace



Εικόνα 2. Πάροχος δεδομένων DSpace με τις επεκτάσεις του Europeana plug-in

4.3 Europeana DSpace plug-in

Το DSpace στην βασική του εγκατάσταση περιλαμβάνει ένα σχήμα βασισμένο στο Qualified Dublin Core. Συνεπώς, οι εγγραφές ενός DSpace αποθετηρίου πρέπει να μετασχηματισθούν ώστε να ακολουθούν το σχήμα μεταδεδομένων της Europeana (ESE, 2009), που επί της ουσίας βασίζεται κι αυτό στο Qualified Dublin Core με προσθήκη ορισμένων πεδίων και προσδιορισμό κάποιων (προαιρετικών) οδηγιών για την κωδικοποίηση τιμών. Τη συγκεκριμένη εργασία διευκολύνει σημαντικά το Europeana DSpace plug-in. Η εικόνα 2

παρουσιάζει (με κίτρινο χρώμα) τα στοιχεία με τα οποία επεκτείνει το plug-in τον default πάροχο δεδομένων του DSpace.

Οι απαιτούμενες ενέργειες για τον προαναφερθέντα μετασχηματισμό μπορούν να διαχωριστούν σε τρία βήματα:

- i. Προσθήκη πρόσθετων πεδίων του ESE στο σχήμα μεταδεδομένων του αποθετηρίου.
- ii. Συμπλήρωση των τιμών στα πρόσθετα πεδία με βάση τιμές πεδίων του σχήματος του default σχήματος του DSpace, σύμφωνα με τις οδηγίες της Europeana.
- iii. Παροχή των εγγραφών μέσω OAI-PMH server.

Για τα βήματα i και ii, το plug-in προσφέρει δύο εναλλακτικές:

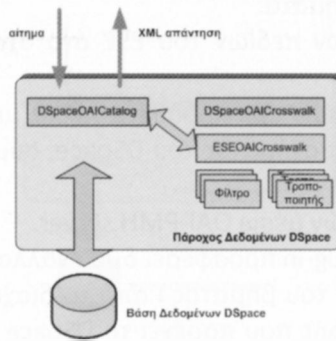
(α) Χειρωνακτική εκτέλεση του βήματος i από το διαχειριστή του αποθετηρίου μέσω της γραφικής διεπαφής που παρέχει το DSpace και χρήση για το βήμα ii του **PHP script** που έχει αναπτυχθεί από τη ΔΚΒΒ.

(β) Χρήση εργαλείου που έχει αναπτυχθεί από το ΕΚΤ σε Java και διεκπεραιώνει αυτόματα τα βήματα i και ii, χωρίς χειρωνακτική παρέμβαση του διαχειριστή του αποθετηρίου. Το εργαλείο αυτό μπορεί να παραμετροποιηθεί μέσω του αρχείου dspace.cfg χωρίς αλλαγή κώδικα στο DSpace. Έχει αναπτυχθεί σε γλώσσα προγραμματισμού Java και χρησιμοποιεί το API του DSpace, οπότε μπορεί να εκτελεστεί εύκολα σε οποιαδήποτε αποθετήριο DSpace χωρίς την εγκατάσταση οποιουδήποτε πρόσθετου λογισμικού ή βιβλιοθήκης.

Το βήμα iii πραγματοποιείται από το **ESECrosswalk** που έχει αναπτυχθεί από τη ΔΚΒΒ, με ενσωματωμένες κάποιες βελτιώσεις από το ΕΚΤ. Το **ESECrosswalk** εκτελείται αντί του DSpaceOAIcCrosswalk που είναι στοιχείο της βασικής εγκατάστασης του DSpace.

4.4 Extended OAI

Το εργαλείο Extended OAI (OAI Extended Addon, 2009) είναι μια προσθήκη στην υποδομή του DSpace που δίνει την δυνατότητα στους διαχειριστές του αποθετηρίου να φιλτράρουν και να μεταβάλλουν με αποδοτικό τρόπο τα μεταδεδομένα που είναι διαθέσιμα προς εξαγωγή στους διάφορους παρόχους υπηρεσιών. Στην εικόνα 3 βλέπουμε την αλληλεπίδραση που έχει το Extended OAI με τον πάροχο υπηρεσιών του DSpace.



Εικόνα 3. Αρχιτεκτονική υλοποίησης Extended OAI Addon.

Με τη βοήθεια του extended OAI, ο διαχειριστής του συστήματος μπορεί να ορίσει εικονικά sets, ανεξάρτητα από την οργάνωση του αποθετηρίου σε κοινότητες / συλλογές. Τα εικονικά sets ορίζονται σε XML αρχείο παραμέτρων του DSpace. Ο πάροχος υπηρεσιών μπορεί να χρησιμοποιήσει τα εικονικά sets σαν ορίσματα του αιτήματος ListRecords του OAI-PMH και το extended OAI είναι υπεύθυνο να επιστρέψει τα κατάλληλα αντικείμενα του αποθετηρίου που ανήκουν στα εικονικά sets με την βοήθεια των φίλτρων (filters) και των τροποποιητών (modifiers). Το περιεχόμενο των εικονικών sets ορίζεται αποκλειστικά από την εφαρμογή κατάλληλων φίλτρων και τροποποιητών.

Το φίλτρο είναι μια μονάδα που έχει ως σκοπό να φιλτράρει τα αντικείμενα προς εξαγωγή στο OAI-PMH, τόσο σε επίπεδο εγγραφών όσο και σε επίπεδο πεδίων. Ο πάροχος δεδομένων του DSpace μπορεί να περιλαμβάνει περισσότερα από ένα φίλτρα ανάλογα των λόγων για τους οποίους κάποια αντικείμενα πρέπει να κοπούν αλλά και το πλήθος των παρόχων υπηρεσιών. Κάθε φίλτρο είναι μια ξεχωριστή κλάση της Java και ενεργοποιείται για ένα ή περισσότερα συγκεκριμένα εικονικά sets.

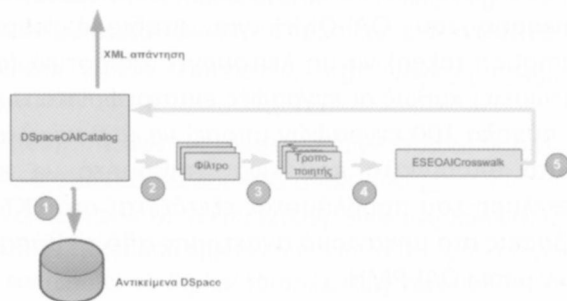
Είναι φανερό πως δεν μπορούν να υπάρχουν καθολικά φίλτρα που να λειτουργούν σε όλα τα αποθετήρια και να φιλτράρουν τις εγγραφές προς εξαγωγή στη Europeana. Και αυτό γιατί κάθε αποθετήριο έχει διαφορετική οργάνωση σε ότι αφορά τις κοινότητες και τις συλλογές αλλά και διαφορετική πολιτική ως προς τα αντικείμενα που πρέπει να διατεθούν στην Europeana. Για το λόγο αυτό, απαιτείται από τη πλευρά των διαχειριστών των αποθετηρίων η ανάπτυξη φίλτρων για το κάθε αποθετήριο χωριστά.

Ο τροποποιητής (*modifier*) είναι μια μονάδα που έχει ως σκοπό να τροποποιήσει τα μεταδεδομένα ενός αντικείμενου. Οι τροποποιητές, καθένας

από τους οποίους ορίζεται σε ξεχωριστή κλάση Java, βρίσκονται σε πλήρη αντιστοιχία με τα φίλτρα ως προς τον τρόπο που δηλώνονται στα αρχεία παραμέτρων αλλά και στο γεγονός ότι μπορούν να οριστούν τα εικονικά sets για τα οποία ισχύει ο κάθε τροποποιητής.

Σε αντίθεση με το Eurorpeana DSpace plug-in, η μέθοδος αυτή του Extended OAI δεν αλλάζει ούτε προσθέτει στοιχεία στη βάση δεδομένων, ουσιαστικά δηλαδή δε χρειάζεται το βήμα i της παραγράφου 4.3. Αντίθετα, σε πραγματικό χρόνο, φιλτράρει τις εγγραφές και τροποποιεί τα μεταδεδομένα και αυτές οι αλλαγές είναι ορατές μόνο για την εκάστοτε συγκομιδή, κάτι που κάνει το μηχανισμό πιο ευέλικτο, καθώς η βάση δεδομένων και το σχήμα του αποθετηρίου δεν επηρεάζεται από τις απαιτήσεις του συσσωρευτή.

Η εικόνα 4 παρουσιάζει τη διαδρομή που ακολουθούν οι εγγραφές του αποθετηρίου πριν εξαχθούν προς συγκομιδή:



Εικόνα 4. Ροή επεξεργασίας εγγραφών εντός του μηχανισμού του Extended OAI.

Μετά τη λήψη του αιτήματος το **DSpaceOAI Catalog** αναζητά τα αντικείμενα προς εξαγωγή. Στη συνέχεια, κάθε αντικείμενο φιλτράρεται από τα διαθέσιμα στο σύστημα φίλτρα, τροποποιούνται τα μεταδεδομένα του από τους διαθέσιμους τροποποιητές και εν τέλει, διαμορφώνεται η XML έξοδος από το **ESEOAI Crosswalk**.

Μπορεί κανείς εύκολα να δει πως το Extended OAI είναι ένα δυνατό εργαλείο στα χέρια των διαχειριστών αποθετηρίων DSpace για τον απόλυτο καθορισμό των μεταδεδομένων που εξάγονται μέσω του πρωτοκόλλου OAI-PMH. Ωστόσο, είναι φανερό πως απαιτεί την ύπαρξη καταρτισμένου προσωπικού πληροφορικής για την εγκατάσταση του συστήματος (απαιτείται εγκατάσταση dspace-oai module διαφορετικού από το βασικό του DSpace) αλλά και την

παραμετροποίησή του με την ανάπτυξη φίλτρων και τροποποιητών, η οποία απαιτεί προγραμματισμό σε Java. Από την άλλη, το Europeana DSpace plug-in έχει λιγότερες διαχειριστικές απαιτήσεις και είναι πολύ ευκολότερο στη χρήση, δεν προσφέρει ωστόσο όλες τις δυνατότητες που παρέχει το extended OAI όσον αφορά το φιλτράρισμα των εγγραφών.

Ένα σημαντικό ζήτημα, η διερεύνηση του οποίου είναι σε εξέλιξη, είναι η αποδοτική ανάκτηση εγγραφών μέσω OAI-PMH όταν (α) ένα αποθετήριο περιέχει σχετικά μεγάλο αριθμό εγγραφών (π.χ. άνω των 10000) και (β) οι εγγραφές προς παροχή σε ένα αίτημα είναι σχετικά λίγες σε σχέση με το σύνολο των εγγραφών του αποθετηρίου. Όταν συμβαίνει κάτι τέτοιο, είναι πιθανό ο μηχανισμός του OAI-OMH για σταδιακή παροχή εγγραφών (μηχανισμός resumption token) να μη λειτουργεί αποδοτικά (ούτε καν σωστά σε κάποιες περιπτώσεις) καθώς οι εγγραφές επιστρέφονται συνήθως ανά 100 και στο εκάστοτε σύνολο 100 εγγραφών μπορεί να μην περιλαμβάνεται καμία που να ανήκει στο υποσύνολο που είναι επιθυμητό να παρασχεθεί στο συσσωρευτή. Η επίλυση του προβλήματος εξετάζεται στο EKT με κατάλληλες δραστικές παρεμβάσεις στο μηχανισμό ανάκτησης από τη βάση δεδομένων και παροχής εγγραφών μέσω OAI-PMH.

Ένα ακόμη θέμα προς μελλοντική εξέταση είναι η υποστήριξη του σχήματος Europeana Data Model (EDM) που θα αποτελέσει το μοντέλο μεταδεδομένων της Europeana σε βάθος χρόνου (EDM, 2010).

5. Σύνοψη – Συμπεράσματα

Το ζήτημα της συμβατότητας των αποθετηρίων με υποδομές συσσωρευτών γίνεται ολοένα και πιο κρίσιμο αλλά και πιο πολύπλοκο στο σύγχρονο περιβάλλον των ψηφιακών βιβλιοθηκών με την αυξανόμενη τάση για δημιουργία συστημάτων συσσωρευτών ευρείας κλίμακας με ιδιαίτερη υποστήριξη από φορείς όπως η Ευρωπαϊκή Ένωση. Το γεγονός αυτό καθιστά την ανάπτυξη και πρακτική χρήση εργαλείων λογισμικού με εξελιγμένες δυνατότητες για την παροχή μεταδεδομένων μέσω OAI-PMH, όπως αυτά που παρουσιάζονται στο παρόν άρθρο. Η εμπειρία στην Ελλάδα δείχνει ότι στη χώρα μας υπάρχει η υποδομή και το κατάλληλο δυναμικό για την επίτευξη διαλειτουργικότητας που θα ωφελήσει σημαντικά τα υπάρχοντα αποθετήρια. Είναι σημαντικό η ελληνική κοινότητα των ψηφιακών βιβλιοθηκών και αποθετηρίων να αξιοποιήσει τις σχετικές ευκαιρίες.

Βιβλιογραφία

- Banos, E. (200). DSpace plugin for Europeana Semantic Elements (ESE), <http://vbanos.gr?p=189> [accessed 25/3/2010]
- Donas. K. (2010). DEIXTo, <http://www.deixto.com> [accessed 25/3/2010]
- Europeana Data Model (2010). Definition of the Europeana Data Model element. Version 5.2, 30/7/2010. http://group.europeana.eu/c/document_library/get_file?uuid=aff89c92-b6ff-4373-a279-fc47b9af3af2&groupId=10605 [Ημερομηνία πρόσβασης: 18-09-2010].
- Europeana Semantic Elements specifications. Version 3.2.1, 06/11/2009. Europeana v1.0 (2009). http://www.europeanfilmgateway.eu/downloads/Europeana_Semantic_Elements_Specifications_v3_2_1_2009-11-06-1.pdf [Ημερομηνία πρόσβασης: 27-06-2010].
- EuropeanaLocal (2008). *EuropeanaLocal: connecting cultural heritage*, <http://www.europeanlocal.eu> [Ημερομηνία πρόσβασης: 27-06-2010].
- Koninklijke Bibliotheek (2009). *Europeana*, <http://www.europeana.eu> [Ημερομηνία πρόσβασης: 27-06-2010].
- Lagoze, C., et al. (2008). *The Open Archives Initiative Protocol for Metadata Harvesting*, <http://www.openarchives.org/OAI/openarchivesprotocol.html> [accessed: 15/8/2009].
- Mazurek, C., M. Mielnicki, and M. Werla (2009). Selective harvesting of regional digital libraries and national metadata aggregators. In *JCDL '09: Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, pp. 429-430. ACM.
- OAI Extended Addon (2009), http://projecto.rcaap.pt/index.php?option=com_remository&Itemid=2&func=startdown&id=71&lang=en, [Ημερομηνία πρόσβασης: 18-09-2010].
- Ouksel A.M. and Sheth A. (1999) Semantic Interoperability in Global Information Systems, ACM SIGMOD Record, Vol 28(1) March 1999, pp 5-12.
- Veria Central Public Library (2010). *Europeana Local Aggregator*, <http://aggregator.libver.gr> [Ημερομηνία πρόσβασης: 27-06-2010]
- Δημόσια Κεντρική Βιβλιοθήκη Βέροιας (2009). *Europeana Local – Δημόσια Κεντρική Βιβλιοθήκη Βέροιας*, <http://www.libver.gr/edlocal/> [Ημερομηνία πρόσβασης: 27-06-2010]
- Δημόσια Κεντρική Βιβλιοθήκη Βέροιας (2010). *Μέδουσα*, <http://medusa.libver.gr/> [Ημερομηνία πρόσβασης: 27-06-2010]