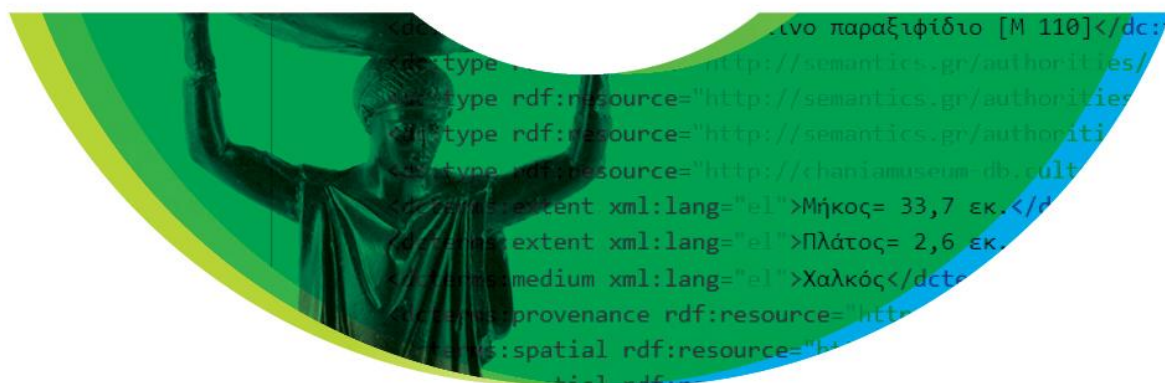




Μια τεχνική ματιά στη συσσώρευση
ψηφιακού περιεχομένου από το ΕΚΤ:
από τη συλλογή & τον έλεγχο
στον εμπλουτισμό & τη διάχυση



{ Search
Culture.gr }

Ενιαία Αναζήτηση
σε ανοικτό ψηφιακό
περιεχόμενο



Μονάδα Ανάπτυξης Εφαρμογών ΕΚΤ

Κείμενο: Χάρης Γεωργιάδης

Σχεδιασμός Έκδοσης: Δήμητρα Πελεκάνου

Επιμέλεια: Έλενα Λαγούδη

Οπτικοποίηση: Χάρης Γεωργιάδης, Δήμητρα Πελεκάνου

Προτεινόμενος τρόπος αναφοράς:

ΕΚΤ(2017), *Οδηγός Χρήσης searchculture.gr, μία πρότυπη πλατφόρμα συσσώρευσης & αναζήτησης πολιτιστικού περιεχομένου*, Αθήνα: Εθνικό Κέντρο Τεκμηρίωσης

Copyright © 2017 Εθνικό Κέντρο Τεκμηρίωσης | ΕΙΕ

δ. Βασιλέως Κωνσταντίνου 48, 11635 Αθήνα | τ.: 210 7273900

| f: 210 7246824 | e: ekt@ekt.gr | www.ekt.gr



Το έργο αυτό διατίθεται με άδεια Creative Commons
Αναφορά-Μη Εμπορική Χρήση-Όχι Παράγωγα Έργα 4.0 Ελλάδα

Προκειμένου να δείτε αντίγραφο της άδειας επισκεφθείτε:

<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.el>

{ Search Culture.gr }

Ενιαία Αναζήτηση σε
ανοικτό ψηφιακό περιεχόμενο



ΠΕΡΙΕΧΟΜΕΝΑ

1	Εισαγωγή	5
2	Οικοσύστημα / συνολική αρχιτεκτονική	6
3	Λειτουργικά χαρακτηριστικά του searchculture.gr.....	10
4	Τεχνικά χαρακτηριστικά του searchculture.gr	13

1 Εισαγωγή

Το Σύστημα Συσσώρευσης Περιεχομένου / Ενιαίας Αναζήτησης searchculture.gr είναι ένας συσσωρευτής μεταδεδομένων και αρχείων προεπισκόπησης (aggregator) ο οποίος συνοδεύεται από μία ανοικτή διαδικτυακή πύλη που προσφέρει κεντρική πρόσβαση και ενιαία αναζήτηση σε ψηφιακό πολιτιστικό περιεχόμενο που παράγεται στη χώρα μας από έγκριτους φορείς. Το ψηφιακό περιεχόμενο συλλέγεται και ελέγχεται ως προς την συμμόρφωσή του με συγκεκριμένες προδιαγραφές διαλειτουργικότητας. Στην συνέχεια, ο συσσωρευτής συγκεντρώνει το περιεχόμενο εκτελώντας ενέργειες μετασχηματισμού, σημασιολογικού εμπλουτισμού, κανονικοποίησης και καθαρισμού, και το διαθέτει στο ευρύ κοινό ως ανοικτά διασυνδεδεμένα δεδομένα μέσω ενός σύγχρονου και φιλικού περιβάλλοντος που διαθέτει προηγμένες λειτουργίες αναζήτησης και εξελιγμένες δυνατότητες πλοήγησης.

Search Culture.gr Ενιαία Αναζήτηση σε ανοικτό ψηφιακό περιεχόμενο

ΕΚΤ ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ NATIONAL DOCUMENTATION CENTRE SaaS Services aggregator

Αρχική σελίδα Αναζήτηση Φορείς και συλλογές ΕΚΤ τύποι τεκμηρίων ΕΚΤ ιστορικές περίοδοι Το έργο Ελληνικά | English

Πληκτρολογήστε λέξη ή φράση Σε όλα τα πεδία Αναζήτηση Καθαρισμός

Αναζήτηση σε όλες τις συλλογές

Επιλέξτε ΕΚΤ τύπους Επιλέξτε άδεια χρήσης ψηφιακού περιεχομένου

Επιλέξτε ΕΚΤ ιστορική περίοδο Ιστορική περίοδος

Αναζήτηση αυστηρώς μέσα στα όρια ιστορικών περιόδων ή χρονολογικών διαστημάτων ON

Ρωμαϊκή Περίοδος Πρώμη Βυζαντινή Περίοδος Μέση Βυζαντινή Περίοδος Ύστερη Βυζαντινή Περίοδος Οθωμανική Περίοδος Νεότερη Ελλάδα

100 μ.Χ. 200 μ.Χ. 300 μ.Χ. 400 μ.Χ. 500 μ.Χ. 600 μ.Χ. 700 μ.Χ. 800 μ.Χ. 900 μ.Χ. 1000 μ.Χ. 1100 μ.Χ. 1200 μ.Χ. 1300 μ.Χ. 1400 μ.Χ. 1500 μ.Χ. 1600 μ.Χ. 1700 μ.Χ. 1800 μ.Χ. 1900 μ.Χ.

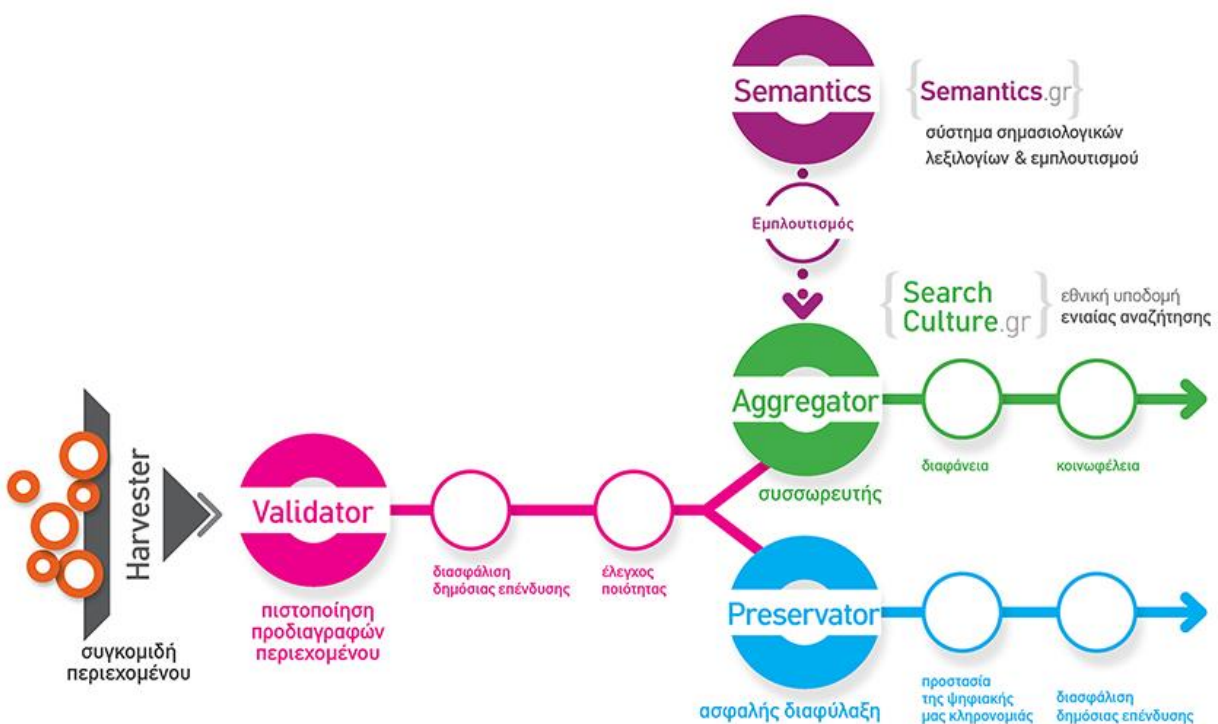
Εξερευνήστε 161881 τεκμήρια ποικίλων τύπων από 43 φορείς περιεχομένου

Αντίγραφο Νομική υπόθεση και σημείωμα Σήμα (σύμβολο) Κεραμικό αντικείμενο Βιβλίο Εκτύπωση Φωτογραφική διαφάνεια Χαρακτικό έργο Ανάγλυφο Αφίσα Αγγείο Αρχαιολογικό αντικείμενο Πρωτότυπο Απόκομμα Άρθρο Ενδυμασία Έγγραφο Σχέδιο/Σκίτσο Επιγραφή Ειδώλιο Γλυπτό Ζωγραφική Χειρόγραφο Τυπωμένο

Η αρχική σελίδα του searchculture.gr

2 Οικοσύστημα / συνολική αρχιτεκτονική

Το searchculture.gr αποτελεί ουσιαστικά τμήμα μιας επεκτάσιμης και παραμετροποιήσιμης πλατφόρμας η οποία αποτελείται από επιμέρους συστήματα και εργαλεία που ολοκληρώνουν το κύκλο ζωής του περιεχομένου από τη συλλογή και τον έλεγχο, μέχρι τον εμπλουτισμό και την διάχυσή του. Η πλατφόρμα χρησιμοποιεί το Σύστημα Συγκομιδής Μεταδεδομένων και Ψηφιακών αρχείων Harvester, το οποίο συλλέγει μεταδεδομένα και ψηφιακά αρχεία από τους φορείς περιεχομένου και επιπροσθέτως δημιουργεί εικόνες προεπισκόπησης. Μετά την συγκομιδή, το περιεχόμενο ελέγχεται και πιστοποιείται από το Σύστημα Πιστοποίησης της Συμμόρφωσης με Προδιαγραφές Διαλειτουργικότητας Validator και εμπλουτίζεται σημασιολογικά από το Σύστημα Λεξιλογίων και Σημασιολογικού Εμπλουτισμού semantics.gr. Η διαδικασία εισαγωγής πραγματοποιείται μέσα από το Διαχειριστικό Περιβάλλον του Συσσωρευτή και το περιεχόμενο τελικά δημοσιοποιείται στη Ανοικτή Διαδικτυακή Πύλη Ενιαίας Αναζήτησης searchculture.gr. Στο τελευταίο βήμα και για τους φορείς που το επιθυμούν, το περιεχόμενο εισάγεται στο Σύστημα Ασφαλούς Διαφύλαξης Περιεχομένου – Preservator. Η πλατφόρμα και τα επιμέρους συστήματα σχεδιάστηκαν και υλοποιήθηκαν εξ ολοκλήρου από το ΕΚΤ. Ακολουθεί μία σύντομη περιγραφή για κάθε ένα από τα επιμέρους συστήματα.



Ο κύκλος ζωής του ψηφιακού περιεχομένου που συσσωρεύεται στις υποδομές του ΕΚΤ



Σύστημα Συγκομιδής Ψηφιακού Περιεχομένου

Το σύστημα Συγκομιδής Ψηφιακού Περιεχομένου **Harvester** συλλέγει μεταδεδομένα και ψηφιακά αρχεία, τα αποθηκεύει και τα διαθέτει κατόπιν αίτησης από χρήστη ή από εξωτερικό σύστημα μέσω της γραφική web διεπαφής χρήστη ή των web υπηρεσιών του (RESTful API), αντίστοιχα. Το περιεχόμενο που συλλέγεται αποθηκεύεται αυτούσιο και μπορεί να επαναχρησιμοποιηθεί πολλές φορές από τρίτα συστήματα όπως τα συστήματα searchculture.gr, Ασφαλούς Διαφύλαξης Ψηφιακού Περιεχομένου και το back-end του συστήματος Validator. Το σύστημα Harvester υποστηρίζει συλλογή μεταδεδομένων και ψηφιακών αρχείων ενώ ειδικά για τα ψηφιακά αρχεία διαθέτει μία υπηρεσία μετατροπής τους σε εικόνες προεπισκόπησης, λειτουργία που αξιοποιείται από τον συσσωρευτή searchculture.gr.



Σύστημα συσσώρευσης περιεχομένου / Πύλη ενιαίας αναζήτησης

Ο Συσσωρευτής/searchculture.gr είναι ένα σύστημα συσσώρευσης (**Aggregator**) το οποίο συσσωρεύει μεταδεδομένα και αρχεία προεπισκόπησης που έχουν συλλεχθεί από το Σύστημα Συγκομιδής Harvester, διεξάγοντας ενέργειες μετασχηματισμού, εμπλουτισμού και κανονικοποίησης, και τα διαθέτει μέσα από μια ανοικτή διαδικτυακή πύλη (www.searchculture.gr). Χάρis στις δυνατότητες σημασιολογικού εμπλουτισμού, καθαρισμού και κανονικοποίησης χρονικής πληροφορίας που υποστηρίζει ο συσσωρευτής, η πύλη searchculture.gr παρέχει προηγμένες δυνατότητες πολυγλωσσικής αναζήτησης και πλοήγησης στο συσσωρευμένο περιεχόμενο που περιλαμβάνουν ιεραρχική πλοήγηση σε τύπους και ιστορικές περιόδους, χρονική αναζήτηση, εφαρμογή φίλτρων αποτελεσμάτων για τύπους, χρονικά διαστήματα και ιστορικές περιόδους, και πλοήγηση μέσω νέφους ετικετών δημοφιλών τύπων και ενός διαδραστικού χρονολογίου. Η πύλη searchculture.gr είναι δίγλωσση και υποστηρίζει πρόσβαση από κινητές συσκευές.

Ο συσσωρευτής διαθέτει ένα φιλικό και λειτουργικό γραφικό διαχειριστικό περιβάλλον μέσω του οποίου ο εξουσιοδοτημένος χρήστης του ΕΚΤ διαχειρίζεται τους φορείς-πηγές, πυροδοτεί εύκολα διαδικασίες εισαγωγής περιεχομένου (μεταδεδομένων και εικόνων προεπισκόπησης), παρακολουθεί την εξέλιξη των διαδικασιών, διαχειρίζεται το συσσωρευμένο περιεχόμενο και πυροδοτεί διαδικασίες σημασιολογικού εμπλουτισμού.



Σύστημα Ελέγχου Συμμόρφωσης με Προδιαγραφές

Το σύστημα Ελέγχου Συμμόρφωσης με Προδιαγραφές Διαλειτουργικότητας **Validator** είναι το πληροφοριακό σύστημα που αναπτύχθηκε από το ΕΚΤ και ελέγχει ψηφιακά αποθετήρια φορέων για το βαθμό κατά τον οποίο συμμορφώνονται με προδιαγραφές ποιότητας και διαλειτουργικότητας μεταδεδομένων, ψηφιακών αρχείων και υπηρεσιών.

Το σύστημα αποτελείται από δύο υποσυστήματα, το back-end, το οποίο είναι το σύστημα που πραγματοποιεί αυτούς καθαυτούς τους ελέγχους και δημιουργεί αναλυτικές τεχνικές αναφορές, και το front-end, μέσα από το οποίο ο

εξουσιοδοτημένος χρήστης πυροδοτεί εύκολα τις διαδικασίες ελέγχου για τα αποθετήρια των εγγεγραμμένων στο σύστημα φορέων, παρακολουθεί τις διαδικασίες και λαμβάνει αναλυτικές αναφορές υψηλού επιπέδου.



Σύστημα Λεξιλογίων και Εργαλείο Σημασιολογικού εμπλουτισμού

Το **semantics.gr** είναι ένα Σύστημα Λεξιλογίων και Σημασιολογικού Εμπλουτισμού που ανέπτυξε το ΕΚΤ όπου το ΕΚΤ αλλά και συνεργαζόμενοι εξουσιοδοτημένοι φορείς καταχωρούν, καθιερώνουν και διασυνδέουν λεξιλόγια σημασιολογικών όρων. Παρέχει μία δημόσια διαδικτυακή πύλη (www.semantics.gr) προσφέροντας ενιαία αναζήτηση σημασιολογικών όρων σε πολλαπλές πηγές και διαθέτει το περιεχόμενο ως Ανοικτά Διασυνδεδεμένα Δεδομένα (Open Linked Data). Το semantics.gr διαθέτει ένα φιλικό στην χρήση μηχανισμό παραμετροποίησης για την μοντελοποίηση των σχημάτων των λεξιλογίων που θα δημιουργούνται από τους φορείς.

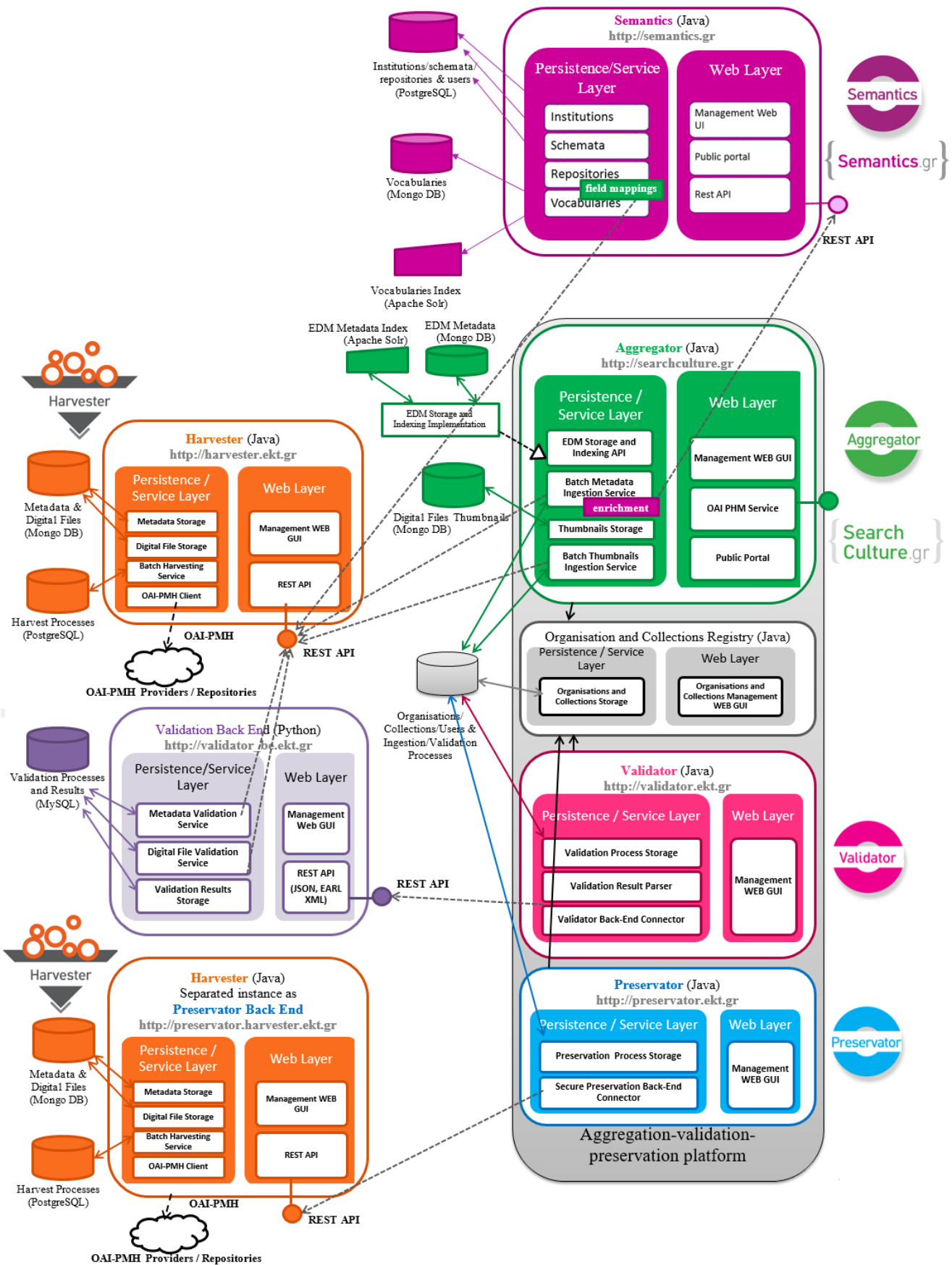
Παράλληλα, το semantics.gr διαθέτει ένα πρωτότυπο και ιδιαίτερα φιλικό εργαλείο για μαζικό ημιαυτόματο σημασιολογικό εμπλουτισμό και ομογενοποίηση περιεχομένου που μπορεί να προσφέρεται ως υπηρεσία τόσο σε αποθετήρια πρωτογενούς περιεχομένου όσο και σε συσσωρευτές. Το ΕΚΤ ήδη χρησιμοποιεί το εργαλείο αυτό προκειμένου να εμπλουτίσει το περιεχόμενο που συσσωρεύεται στο searchculture.gr με συνδέσμους σε καθιερωμένα λεξιλόγια και θησαυρούς όρων αλλά και για να ομογενοποιήσει το περιεχόμενο αυτό με στόχο την βελτιστοποίηση των υπηρεσιών αναζήτησης και πλοήγησης της διαδικτυακής πύλης του searchculture.gr.



Υπηρεσία ασφαλούς διαφύλαξης

Το σύστημα Ασφαλούς Διαφύλαξης Ψηφιακού Περιεχομένου **Preservator** στόχο έχει την επαύξηση της διαθεσιμότητας και της ακεραιότητας των Ψηφιακών Πολιτιστικών Πόρων των αποθετηρίων των φορέων, σύμφωνα με τις αρχές που θέτουν διεθνή πρότυπα για την ασφαλή παροχή και λειτουργία των Πληροφοριακών Συστημάτων. Αφορά απομακρυσμένη, από το κύριο σημείο διάθεσης, ασφαλή αποθήκευση των Ψηφιακών Πολιτιστικών Πόρων και μηχανισμούς διασφάλισης της διαθεσιμότητας και ακεραιότητας τους σε βάθος χρόνου, σεβόμενη τις απαιτήσεις ελέγχου πρόσβασης.

Το σύστημα Preservator αποτελείται από δύο υποσυστήματα, το back-end), το οποίο είναι ουσιαστικά μία ξεχωριστή εγκατάσταση του Συστήματος Συγκομιδής Harvester και το front-end μέσα από το οποίο ο εξουσιοδοτημένος χρήστης πυροδοτεί εύκολα τις διαδικασίες συλλογής περιεχομένου των αποθετηρίων των εγγεγραμμένων στο σύστημα φορέων και τις διαδικασίες επανάκτησης του περιεχομένου κατόπιν αίτησης του φορέα.



Συνολική αρχιτεκτονική των συστημάτων του ΕΚΤ για την συλλογή, πιστοποίηση, εμπλουτισμό, διάχυση και ασφαλή διαφύλαξη ψηφιακού περιεχομένου

3 Λειτουργικά χαρακτηριστικά του Συσσωρευτή searchculture.gr

Ο συσσωρευτής searchculture.gr αποθηκεύει τα μεταδεδομένα που συσσωρεύει σε μορφή συμβατή με τον τύπο μεταδεδομένων EDM¹. Το EDM είναι το μοντέλο που προτείνεται από την Europeana για την αναπαράσταση μεταδεδομένων πολιτιστικής κληρονομιάς και αποτελεί μετεξέλιξη του ESE². Το EDM είναι ένα RDF μοντέλο και ενσωματώνει μηχανισμούς που επιτρέπουν στα μεταδεδομένα να διατηρούν σημασιολογικές αναφορές σε λεξιλόγια και θησαυρούς όρων καθιστώντας εφικτή την διάθεσή τους ως Διασυνδεδεμένα Δεδομένα. Επομένως, το searchculture.gr μπορεί να φιλοξενήσει εγγενώς σημασιολογικά εμπλουτισμένο περιεχόμενο. Εσωτερικά η Europeana χρησιμοποιεί μια παραλλαγή του δημόσιου EDM τύπου, γνωστή ως EDM-internal, η οποία είναι προσαρμοσμένη για τις ανάγκες ενός συσσωρευτή. Το searchculture.gr, ως συσσωρευτής, αποθηκεύει τα μεταδεδομένα στο εσωτερικό EDM σχήμα.

Ο συσσωρευτής υποστηρίζει εισαγωγή μεταδεδομένων σε 4 εναλλακτικούς τύπους. Αποθετήρια τα οποία δεν περιέχουν σημασιολογικά εμπλουτισμένο περιεχόμενο μπορούν να διαθέσουν τα μεταδεδομένα τους σε τύπο OAI-DC ή ESE. Τα αποθετήρια που διαθέτουν σημασιολογικά εμπλουτισμένο περιεχόμενο μπορούν να διαθέσουν τα μεταδεδομένα τους σε EDM (στην δημόσια έκδοσή του). Εναλλακτικά, μπορούν να διαθέσουν τα μεταδεδομένα τους συνδυαστικά σε δύο τύπους, σε ESE και σε DC-DS-XML³. Ο τύπος DC-DS-XML είναι μία XML αναπαράσταση του Dublin Core (Metadata Terms) με δυνατότητα σημασιολογικών αναφορών σε όρους λεξιλογίων και θησαυρών. Στην περίπτωση αυτή, οι πληροφορίες για τα ψηφιακά αρχεία αντλούνται από τα μεταδεδομένα σε τύπο ESE. Η βασική ροή εισαγωγής μεταδεδομένων του searchculture.gr μετασχηματίζει τα πρωτότυπα μεταδεδομένα από EDM, OAI-DC, ESE ή DC-DS-XML στον τύπο EDM-internal προκειμένου να τα αποθηκεύσει στην βάση δεδομένων που χρησιμοποιεί.

Το searchculture.gr παρέχει υψηλού επιπέδου υπηρεσίες αναζήτησης, πλοήγησης και παρουσίασης πολιτιστικού περιεχομένου προβάλλοντας και αναδεικνύοντας με τον καλύτερο τρόπο το ψηφιακό περιεχόμενο έγκριτων φορέων πολιτισμού. Βασικοί στόχοι του searchculture.gr είναι τόσο η αύξηση της ευρεσιμότητας όσο και η αύξηση του όγκου του ετερόκλητου περιεχομένου που συσσωρεύει. Οι δύο αυτοί στόχοι είναι κοινói για όλους τους συσσωρευτές περιεχομένου και η ικανοποίηση και των δύο αποτελεί πρόκληση καθώς οι λειτουργικές απαιτήσεις που αυτοί συνεπάγονται δρουν ανταγωνιστικά μεταξύ τους, όπως φαίνεται παραστατικά στο ακόλουθο σχήμα.

¹EDM documentation, <http://pro.europeana.eu/page/edm-documentation>

²ESE documentation, <http://pro.europeana.eu/page/ese-documentation>

³Expressing Dublin Core Description Sets using XML (DC-DS-XML), available at <http://dublincore.org/documents/dc-ds-xml/>, date issued: 2008-09-01



Οι δύο βασικοί στόχοι ενός συσσωρευτή και ενδεικτικοί τρόποι ικανοποίησής τους

Το searchculture.gr καταφέρνει να ικανοποιήσει και τους δύο αυτούς στόχους σε πολύ ικανοποιητικό βαθμό, χάρις στις υπηρεσίες προστιθέμενης αξίας που παρέχει το ΕΚΤ (δημοσιοποίηση προδιαγραφών διαλειτουργικότητας, έλεγχος/πιστοποίηση, εμπλουτισμός περιεχομένου) αλλά και στα ποιοτικά συστήματα που ανέπτυξε τα οποία καθιστούν εφικτή την υλοποίηση των υπηρεσιών αυτών καθώς και την διαχείριση όλου αυτού του όγκου πληροφορίας (αυτοματοποίηση διαδικασιών ελέγχου, εισαγωγής και εμπλουτισμού, έξυπνος τρόπος μαζικής σημασιολογικής ομογενοποίησης, εξελιγμένο διαχειριστικό περιβάλλον, προηγμένες δυνατότητες αναζήτησης και πλοήγησης περιεχομένου).

Το ΕΚΤ διεξάγοντας αυτοματοποιημένους ελέγχους ποιότητας στο περιεχόμενο που συσσωρεύεται με το σύστημα Validator εξασφαλίζει, σε συνεργασία με τους φορείς, την υψηλή ποιότητα του περιεχομένου που εισάγεται στο searchculture.gr. Η ποιότητα της τεκμηρίωσης βελτιώνει την παρουσίαση, την ευρεσιμότητα και διευκολύνει την αναζήτηση του περιεχομένου και κατ' επέκταση αυξάνει την συνολική ποιότητα των υπηρεσιών που προσφέρει η διαδικτυακή πύλη searchculture.gr στους πολίτες και στους φορείς.

Παράλληλα, η ποιότητα υπηρεσιών αυξάνεται περαιτέρω χάρις στον σημασιολογικό εμπλουτισμό και την ομογενοποίηση του περιεχομένου που πραγματοποιεί το ΕΚΤ με τη βοήθεια του Συστήματος Λεξιλογίων semantics.gr αλλά και του ενσωματωμένου στο συσσωρευτή εργαλείου κανονικοποίησης χρονικής πληροφορίας. Τα τεκμήρια εμπλουτίζονται με ένα νέο πεδίο «ΕΚΤ Τύπος» που περιλαμβάνει αναφορές στους όρους ενός δίγλωσσου ιεραρχικού λεξιλογίου τύπων που δημιούργησε το ΕΚΤ. Επιπλέον, το ΕΚΤ εμπλουτίζει το περιεχόμενο που συσσωρεύει με ομογενοποιημένη χρονολόγηση και το ταξινομεί σε ιστορικές περιόδους βάσει ενός πρότυπου δίγλωσσου ιεραρχικού λεξιλογίου ιστορικών περιόδων. Τα τεκμήρια εμπλουτίζονται με δύο νέα χρονικά πεδία, «ΕΚΤ Χρονολόγηση» και «ΕΚΤ Ιστορική

Περίοδος». Τόσο το λεξιλόγιο τύπων όσο και το λεξιλόγιο ιστορικών περιόδων δημιουργήθηκαν στο semantics.gr. Το αποτέλεσμα του σημασιολογικού εμπλουτισμού επέτρεψε την ανάπτυξη προηγμένης πολυγλωσσικής αναζήτησης και πλοήγησης, συμπεριλαμβανομένης της ιεραρχικής πλοήγησης σε τύπους και ιστορικές περιόδους, της εφαρμογής φίλτρων αποτελεσμάτων για τύπους, χρονικά διαστήματα και ιστορικές περιόδους, και της πλοήγησης μέσω νέφους ετικετών και ενός διαδραστικού χρονολογίου. Τα οφέλη συνοψίζονται ως εξής:

- Αυξάνεται η ευρεσιμότητα του περιεχομένου αφού ο χρήστης μπορεί να ανακαλύψει περιεχόμενο εξερευνώντας μέσα από ένα ομοιόμορφο και συμπαγές σύνολο τύπων ή ιστορικών περιόδων ακόμα και αν δεν γνωρίζει ακριβώς τι ψάχνει.
- Δίνεται η δυνατότητα σε αγγλόφωνο κοινό να ανακαλύψει τεκμήρια βάσει του τύπου τους ή της ιστορικής περιόδου στην οποία χρονολογούνται ακόμα και αν τα μεταδεδομένα τους πρωτοτύπως δεν περιείχαν καθόλου πολυγλωσσική πληροφορία.
- Δίνεται η δυνατότητα χρονικής αναζήτησης και φιλτραρίσματος, τόσο με ιστορικές περιόδους όσο και με χρονολογικά διαστήματα, δυνατότητα η οποία είναι ιδιαίτερα χρήσιμη για πολιτιστικό και ιστορικό περιεχόμενο.
- Συλλογές που είναι σημασιολογικά εμπλουτισμένες ομογενοποιούνται με στόχο την ποιοτικότερη πολυγλωσσική ενιαία αναζήτηση και εξερεύνηση.
- Συλλογές που δεν περιλαμβάνουν καθόλου σημασιολογικές αναφορές μπορούν πλέον να εμπλουτιστούν και να αναδειχθούν μέσα από το searchculture.gr ως Διασυνδεδεμένα Δεδομένα.

Η διαδικτυακή πύλη searchculture.gr προσφέρει προηγμένες δυνατότητες αναζήτησης, οργισμένες πρωτότυπες για συστήματα συσσώρευσης, οι οποίες προσεγγίζουν αυτές που προσφέρονται συνήθως από συστήματα απόθεσης πρωτογενούς περιεχομένου. Οι δυνατότητες περιλαμβάνουν:

- Αναζήτησης βάσει λέξης ή φράσης με λειτουργία προτάσεων αυτόματης συμπλήρωσης όρων (autocomplete) και δυνατότητα περιορισμού της αναζήτησης σε συγκεκριμένα πεδία, όπως ο πρωτότυπος τύπος, ο Europeana τύπος, ο δημιουργός, η θεματική κατηγορία και γεωγραφική κάλυψη
- Σύνθετη αναζήτηση
- Δυνατότητα αναζήτησης βάσει συλλογής (λίστα επιλογής), EKT τύπου (ιεραρχική λίστα επιλογής), EKT ιστορικής περιόδου (ιεραρχική λίστα επιλογής) και χρονολογικού διαστήματος
- Φίλτρα αναζήτησης στα αποτελέσματα βάσει μιας μεγάλης γκάμας βασικών πεδίων που περιλαμβάνει τον EKT τύπο τεκμηρίου, την EKT ιστορική περίοδο, τον πρωτότυπο τύπο, την θεματική κατηγορία, την άδειας χρήσης, τον Europeana τύπο και την συλλογή. Έτσι ο χρήστης μπορεί να εντοπίσει πιο εύκολα τα τεκμήρια που αναζητά φιλτράροντας συνδυαστικά τα αποτελέσματα και μάλιστα χρησιμοποιώντας ως φίλτρα, όχι μόνο τεχνικά πεδία ή πεδία που σχετίζονται με την διάθεση του τεκμηρίου (όπως τα δικαιώματα χρήσης και ο πάροχος), αλλά και πεδία που ορίζουν και περιγράφουν σημασιολογικά το ίδιο το τεκμήριο και που βοηθούν καταλυτικά στην ανακάλυψή του (όπως ο EKT τύπος, η ιστορική περίοδος, ο πρωτότυπος τύπος και η θεματική κατηγορία).

Εκτός από την μηχανή αναζήτησης, η διαδικτυακή πύλη searchculture.gr προσφέρει προηγμένες δυνατότητες δίγλωσσης πλοήγησης μέσα από την αρχική σελίδα οι οποίες περιλαμβάνουν

- Διαδραστικό νέφος ετικετών δημοφιλών ομογενοποιημένων EKT τύπων τεκμηρίων

- Διαδραστικό χρονολόγιο
- Διαδραστική πίτα με τις μεγαλύτερες συλλογές που φιλοξενούνται από το searchculture.gr
- Ξεχωριστή σελίδα ιεραρχικής πλοήγησης μέσα από τους ΕΚΤ τύπους
- Ξεχωριστή σελίδα ιεραρχικής πλοήγησης μέσα από τους ΕΚΤ ιστορικές περιόδους
- Ξεχωριστή σελίδα παρουσίασης των φορέων και πλοήγησης στις συλλογές τους

Παράλληλα, το SearchCulure.gr αποτελεί μία μοντέρνα και εύχρηστη διαδικτυακή πύλη με φιλική και ελκυστική διεπαφή χρήστη, καθώς:

- Υποστηρίζει δίγλωσσο περιβάλλον σε Ελληνικά και Αγγλικά σε στατικό και δυναμικό περιεχόμενο
- Όλα σχεδόν τα τεκμήρια που φιλοξενούνται συνοδεύονται από εικόνες προεπισκόπησης κάνοντας την αναζήτηση πιο ελκυστική αλλά και πιο λειτουργική.
- Οι σελίδες ανταποκρίνονται και προσαρμόζονται ανάλογα με την ανάλυση της οθόνης του χρήστη. Έτσι το searchculture.gr είναι προσβάσιμο το ίδιο εύχρηστα και λειτουργικά και μέσα από φορητές συσκευές (κινητά τηλέφωνα, tablets)

Τέλος, πέραν των ποιοτικών υπηρεσιών που προσφέρει η διαδικτυακή πύλη του searchculture.gr στους πολίτες και στους φορείς περιεχομένου, αξίζει να τονιστεί ότι το σύστημα διαθέτει ένα πλήρες γραφικό διαχειριστικό περιβάλλον το οποίο επιτρέπει στο προσωπικό του ΕΚΤ να διαχειριστεί εύκολα, μαζικά και αποδοτικά πολύ μεγάλο όγκο δεδομένων προερχόμενα από μεγάλο αριθμό αποθετηρίων/συλλογών. Συγκεκριμένα, το διαχειριστικό περιβάλλον επιτρέπει στον εξουσιοδοτημένο χρήστη να εκτελεί εύκολα και όσο το δυνατόν πιο αυτοματοποιημένα όλες τις απαραίτητες διαδικασίες για την λειτουργία του συσσωρευτή, όπως, 1) να διαχειριστεί πληροφορίες για φορείς και τις συλλογές/αποθετήριά τους, 2) να παραμετροποιήσει δυναμικά την διαδικτυακή πύλη 2) να παραμετροποιήσει το εργαλείο κανονικοποίησης χρονικής πληροφορίας 4) να ορίσει παραμετρικά ενέργειες μετασχηματισμού, καθαρισμού, εμπλουτισμού και κανονικοποίησης μεταδεδομένων ανά συλλογή 5) να πυροδοτήσει με απόλυτα ελεγχόμενο τρόπο νέες διαδικασίες συλλογής και εισαγωγής μεταδεδομένων ή αρχείων προεπισκόπησης, 6) να διαχειρίζεται μαζικά το περιεχόμενο ανά συλλογή.

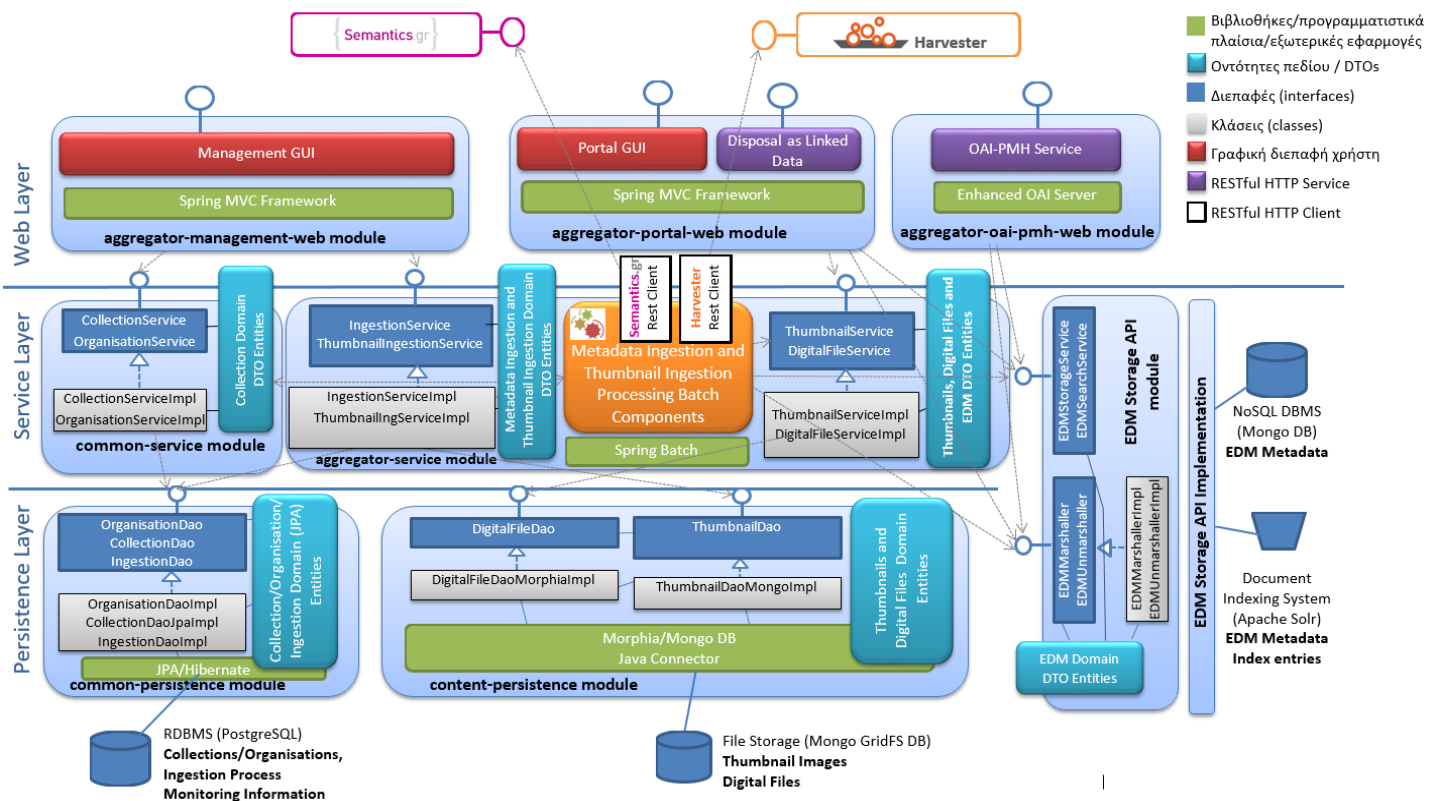
4 Τεχνικά χαρακτηριστικά του searchculture.gr

Το Σύστημα Συσσωρευτή /Ενιαίας Αναζήτησης searchculture.gr είναι μία Java εφαρμογή ιστού πακεταρισμένη ως τρία ξεχωριστά war αρχεία (διαδικτυακή πύλη, διαχειριστικό περιβάλλον, ΟΑΙ-ΡΜΗ υπηρεσία) και μπορεί να εγκατασταθεί σε οποιονδήποτε Servlet container, όπως ο Apache Tomcat. Το σύστημα σχεδιάστηκε σύμφωνα με τις αρχές της πολυ-διαστρωματωμένης αρχιτεκτονικής (multi-layered architecture) και χωρίζεται στα στρώματα persistence (ή αλλιώς infrastructure layer), service (ή αλλιώς business layer) και presentation (εν προκειμένου web layer).

Το persistence layer παρέχει τις βασικές CRUD⁴ μεθόδους καθώς και μεθόδους αναζήτησης για τις οντότητες πεδίου που αποθηκεύονται μόνιμα σε βάσεις δεδομένων. Το searchculture.gr χρησιμοποιεί άμεσα τρεις βάσεις δεδομένων και ένα σύστημα ευρετηρίασης:

⁴ Create, Read, Update, Delete, https://en.wikipedia.org/wiki/Create,_read,_update_and_delete

1. Μία σχεσιακή βάση (PostgreSQL⁵) για την αποθήκευση πληροφορίας σχετικά με τους φορείς/παρόχους και τις συλλογές τους, τις διαδικασίες εισαγωγής μεταδεδομένων και εικόνων προεπισκόπησης, την καταγραφή σφαλμάτων κατά την εισαγωγή και την παραμετροποίηση του συστήματος.
2. Μία noSQL Mongo DB⁶ βάση δεδομένων για την αποθήκευση EDM μεταδεδομένων.
3. Ένα σύστημα ευρετηρίασης Apache Solr⁷, για την ευρετηρίαση των EDM μεταδεδομένων.
4. Ένα File Storage (Mongo DB GridFS⁸) για την αποθήκευση εικόνων προεπισκόπησης και ψηφιακών αρχείων.



Αρχιτεκτονική του συστήματος συσσωρευτή / ενιαίας αναζήτησης searchculture.gr

Στο persistent layer, για όλες τις βάσεις δεδομένων που χρησιμοποιούνται, ορίζονται οι οντότητες πεδίου ως Java POJOs οι οποίες αντιστοιχίζονται με τις αντίστοιχες δομές (PostgreSQL πίνακες ή Mongo DB collections) μέσω ειδικών προγραμματιστικών προτύπων/βιβλιοθηκών (JPA ή Morphia⁹). Για κάθε

⁵ <https://www.postgresql.org/>

⁶ <https://www.mongodb.com/>

⁷ <http://lucene.apache.org/solr/>

⁸ <http://docs.mongodb.org/manual/core/gridfs/>

⁹ <https://github.com/mongodb/morphia/wiki>

βασική οντότητα πεδίου ορίζεται από ένα αντικείμενο με CRUD μεθόδους και μεθόδους αναζήτησης, σύμφωνα με το σχεδιαστικό πρότυπο DAO¹⁰.

Το service layer περιλαμβάνει modules που παρέχουν μία σειρά από υπηρεσίες (services) στα modules του επόμενου layer οι οποίες υλοποιούν όλες τις business διαδικασίες του συσσωρευτή: εκκίνηση νέας διαδικασίας εισαγωγής, REST κλίσεις στο σύστημα Συγκομιδής Harvester, REST κλίσεις στο σύστημα λεξιλογίων και σημασιολογικού εμπλουτισμού semantics.gr, παρακολούθηση και καταγραφή της προόδου της κάθε διαδικασίας, διάθεση αποθηκευμένων εγγραφών μεταδεδομένων, αναζήτηση περιεχομένου κα. Για να επιτευχθεί η αποδέσμευση του επόμενου layer από το persistent layer, άρα και από τις οντότητες πεδίου που αυτό ορίζει, δημιουργήθηκαν Data Access Services για κάθε αντικείμενο DAO και υιοθετήθηκε αυστηρά το DTO προγραμματιστικό υπόδειγμα¹¹, δημιουργήθηκαν δηλαδή νέες οντότητες πεδίου στο Service Layer, τα DTOs. Τα services χρησιμοποιούν ως είσοδο ή/και επιστρέφουν αποκλειστικά DTO αντικείμενα.

Το web layer υλοποιεί την διεπαφή ιστού. Αποτελείται από 1) την δημόσια διαδικτυακή πύλη (Portal) 2) το διαχειριστικό περιβάλλον (management GUI) που επιτρέπει σε έναν εξουσιοδοτημένο χρήστη από το προσωπικό του ΕΚΤ να εκτελεί όλες τις διαχειριστικές λειτουργίες του συσσωρευτή, και 3) την υπηρεσία ιστού OAI-PMH μέσω της οποίας τα μεταδεδομένα διατίθενται για συγκομιδή σε τρίτα συστήματα (όπως ο συσσωρευτής Europeana).

Όλα τα modules χρησιμοποιούν το Spring framework ¹²για Dependency Injection και για την διαχείριση των διεπαφών (interfaces) και των κλάσεων που τις υλοποιούν. Για τον συντονισμό και ενορχήστρωση των κλάσεων που αναπτύχθηκαν και που υλοποιούν τις ροές συγκομιδής/εισαγωγής μεταδεδομένων και εικόνων προεπισκόπησης χρησιμοποιήθηκε το Spring Batch Framework¹³. Για την ανάπτυξη της εφαρμογής ιστού χρησιμοποιήθηκε το Spring MVC framework σε συνδυασμό με front-end τεχνολογίες όπως JQuery¹⁴ και Bootstrap¹⁵. Για την ανάπτυξη της OAI-PMH υπηρεσίας ιστού χρησιμοποιήθηκε η βιβλιοθήκη Enhanced OAI Server ¹⁶ που ανέπτυξε το ΕΚΤ και που έχει χρησιμοποιηθεί ήδη σε πληθώρα αποθετηρίων που δημιούργησε το ΕΚΤ για την διάθεση του περιεχομένου τους για συγκομιδή.

¹⁰Data access object Pattern, http://en.wikipedia.org/wiki/Data_access_object

¹¹Data Transfer Object (DTO), <http://martinfowler.com/eaCatalog/dataTransferObject.html>

¹² <http://spring.io/>

¹³ <http://docs.spring.io/spring-batch/>

¹⁴ The JQuery JS framework, <https://jquery.com/>

¹⁵ Bootstrap, HTML, CSS, and JS framework <http://getbootstrap.com/>

¹⁶ EKT / Enhance OAI-PMH Service, <https://github.com/EKT/EnhancedOAIserver>



Ενιαία Αναζήτηση
σε ανοικτό ψηφιακό περιεχόμενο



ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ
www.ekt.gr