

Σχεδιασμός του Ολοκληρωμένου Συστήματος Ψηφιακής Βιβλιοθήκης του Πανεπιστημίου Αθηνών

Κώστας Βίγλας, Ειρήνη Λουρδή, Μάρα Νικολαΐδη,
Γιώργος Πυρουνάκης, Κώστας Σαΐδης



Πανεπιστήμιο Αθηνών

Υπολογιστικό Κέντρο Βιβλιοθηκών



Περιεχόμενα

- Πώς οδηγούμαστε στο σχεδιασμό της αρχιτεκτονικής
 - Στόχοι / Πλαίσιο ενεργειών
 - Γενική Περιγραφή (Συλλογές, Ανάγκες, Προβλήματα)
 - Στόχοι, Απαιτήσεις, Προδιαγραφές και Επιλογή ΣΨΒ
 - Τα ΣΨΒ Fedora, DSpace, Greenstone
- Παρουσίαση της αρχιτεκτονικής
- Τρέχουσα κατάσταση και μελλοντικά σχέδια



Βασικό Πλαίσιο - Στόχοι

- Δημιουργία, Συντήρηση και Πρόσβαση σε Ψηφιακές Συλλογές ιδιαίτερης πολιτιστικής και εθνικής αξίας (στο πλαίσιο έργων του Β' ΕΠΕΑΕΚ και της ΚτΠ).
- Σχεδιασμός και υλοποίηση ενός ανοικτού και παραμετροποιήσιμου Συστήματος Ψηφιακής Βιβλιοθήκης (ΣΨΒ) που να επιτρέπει με εύκολο, ασφαλή και οικονομικό τρόπο την ενσωμάτωση ετερογενών ψηφιακών συλλογών από το Πανεπιστήμιο Αθηνών.
- Ο σχεδιασμός, η υλοποίηση και η υποστήριξη γίνεται από το Υπολογιστικό Κέντρο Βιβλιοθηκών (ΥΚΒ) του Πανεπιστημίου Αθηνών.



Ψηφιακές Συλλογές

- Λαογραφική Συλλογή
- Ιατρικές Εικόνες
- Μουσειακό Υλικό Ορυκτολογίας - Παλαιοντολογίας
- Αρχείο Θεατρικών Παραστάσεων
- Αρχαίοι πάπυροι
- Ερευνητικό υλικό – λογισμικό
- Ενημερωτικό υλικό
- Μεταπτυχιακές Εργασίες, Διδακτορικές Διατριβές
- Ιστορικό Αρχείο Παν/μίου
- Συλλογή Μουσικών Κωδίκων Βιβ/κης Ψάχου

Χαρακτηριστικά Συλλογών

Τα χαρακτηριστικά των συλλογών ορίζουν προδιαγραφές για το ΣΨΒ

Μεγάλη ανομοιογένεια

- Ποικιλία μορφής
 - Φωτογραφίες, Ψηφιοποιημένο και Ψηφιακό Κείμενο, Έντυπα και Χειρόγραφα, Ήχος, Χάρτες, διάφορα τρισδιάστατα αντικείμενα
 - Υλικό κυρίως ψηφιοποιημένο και όχι αμιγώς ψηφιακό
- Ποικιλία Θεματολογίας
 - Πώς χαρακτηρίζουμε το υλικό;
 - Χρειαζόμαστε διαφορετικά πρότυπα μεταδεδομένων;
- Ποικιλία και Πολυπλοκότητα Δομής
 - Ιεραρχίες, συσχετίσεις
- Απευθύνονται σε διαφορετικό κοινό
 - Πώς παρουσιάζουμε το υλικό;
- Διαφοροποιούνται σε φύση και αξία
 - Χρειαζόμαστε διαφορετικές πολιτικές πρόσβασης



Λαογραφική Συλλογή (I)

- Η πιο **σύνθετη** συλλογή του ΕΚΠΑ
 - Χειρόγραφα τετράδια με φωτογραφίες, χάρτες, σχήματα, εικόνες, δείγματα ενδυμάτων & φυτών κ.λ.π.
- Αποτελεί **συλλογή-πρότυπο** για το σχεδιασμό του Συστήματος
 - Ουσιαστικά θεωρείται ένα υπερσύνολο των υπολοίπων συλλογών



Λαογραφική Συλλογή (II)

- Εμπεριέχει υποσυλλογές
- Τα αντικείμενά της είναι σύνθετης μορφής (αποτελούνται από επιμέρους τμήματα, π.χ. κείμενο και εικόνα μαζί)
- Ορίζει συσχετίσεις μεταξύ αντικειμένων (δομή ιεραρχίας)
- Απαιτεί την παράλληλη χρήση πολλαπλών σχημάτων μεταδεδομένων για την περιγραφή της σε διάφορα επίπεδα χρήσης



Βασικές Ανάγκες για το ΣΨΒ του ΕΚΠΑ

- Διατήρηση του υλικού
 - Αποθήκευση του υλικού και των σχετικών μεταδεδομένων
- Αποδοτική Χρήση του υλικού
 - Εύκολη και αποτελεσματική / χρηστική πρόσβαση μέσω Διαδικτύου
- Υποστήριξη διαλειτουργικότητας και ολοκλήρωσης
- Αποδοτικότερη διαχείριση με ενιαίο τρόπο



Προβλήματα

- Μπορούμε να καλύψουμε τις ανάγκες με ενιαίο τρόπο;
- Πώς μπορούμε να εμπλουτίσουμε τη πρόσβαση στο υλικό – δεν υπάρχει άμεσα η δυνατότητα αναζήτησης στο περιεχόμενο
- Επεκτασιμότητα / Ευελιξία – προσθήκη νέων συλλογών με τον απλούστερο δυνατό τρόπο



Στόχοι για την υλοποίηση του ΣΨΒ

- Ανάπτυξη ενιαίου ΣΨΒ με κοινή υποδομή και παραμετροποιήσιμη λειτουργικότητα ανά συλλογή
- Εισαγωγή όλων των συλλογών σε αυτό
 - Ενιαία διεπαφή χρήσης, προσαρμόσιμη όμως σε κάθε σύλλογή
- Δυνατότητες ενιαίας αναζήτησης σε όλες τις Συλλογές (metadata crosswalks)



Προδιαγραφές ΣΨΒ (I)

- Υποστήριξη πολλαπλών ψηφιακών συλλογών με ενιαία διεπαφή χρήστη
- Υποστήριξη πολλαπλών προτύπων μεταδεδομένων
 - Ως βάση να χρησιμοποιείται το Dublin Core και οι επεκτάσεις του
- Υποστήριξη διαλειτουργικότητας μέσω προτύπων όπως του Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) για την ανταλλαγή των μεταδεδομένων.



Προδιαγραφές ΣΨΒ (II)

- Δυνατότητα αποθήκευσης του υλικού σε πολλές διαφορετικές μορφές
- Δυνατότητα χαρακτηρισμού των συλλογών σε διαφορετικά επίπεδα βάθους και λεπτομέρειας (περιγραφή, χρήση, πνευματικά δικαιώματα, παρουσίαση)
- Υποστήριξη Ροής Εργασίας για τη διαδικασία Χαρακτηρισμού



Προδιαγραφές ΣΨΒ (III)

- Υποστήριξη προηγμένου συστήματος authentication / authorization (χρήστες και ομάδες με διαφορετικά δικαιώματα & αναγνώριση IP διεύθυνσης)
- Υποστήριξη πολλαπλών γλωσσών (τουλάχιστον Ελληνικής και Αγγλικής) σε επίπεδο ψηφιακού υλικού, μεταδεδομένων και παρουσίασης
- Ευελιξία και επεκτασιμότητα
- Υποστήριξη επιπλέον λειτουργικότητας και παραμετροποίησης με προκαθορισμένο τρόπο



Επιλογή ΣΨΒ

- Δεν υπάρχει σύστημα που να καλύπτει τις ανάγκες του ΕΚΠΑ “out of the box”
- Επενδύουμε σε κάποιο σύστημα Λογισμικού Ανοιχτού Κώδικα και το επεκτείνουμε
- Το σύστημα πρέπει να βασίζεται σε στρωματοποιημένη και ανοιχτή αρχιτεκτονική και να είναι συμβατό με την υπάρχουσα υποδομή του ΥΚΒ
- Έμφαση σε επεκτασιμότητα και ευελιξία



Διερεύνηση ΣΨΒ

- Σύγκριση ΣΨΒ Ανοιχτού Κώδικα (βασιζόμενοι και στις προτάσεις του Open Society Institute):
 - **Fedora** (Cornell & Virginia University)
 - **Dspace** (MIT & HP)
 - **Greenstone** (Waikato University, Unesco)
- Διερεύνηση ανάπτυξης της Λαογραφικής Συλλογής με τα παραπάνω συστήματα

The logo consists of a vertical black line on the left, a horizontal black line at the bottom, and three overlapping squares: a yellow one at the top left, a red one at the middle left, and a blue one at the bottom left. The word "Greenstone" is written in a blue serif font to the right of the graphic.

Greenstone

- Ένα αρκετά διαδεδομένο σύστημα με βασική οντότητα την συλλογή
- Χρησιμοποιεί το DC ως πρότυπο μεταδεδομένων
- Αποθηκεύει το υλικό του στην εσωτερική μορφή του Greenstone Archive Format
- Διατίθεται υπό την GNU Public License
- Κρίθηκε ανεπαρκές για τις ψηφιακές συλλογές του ΕΚΠΑ:
 - Δεν προβλέπει λειτουργίες διατηρησιμότητας υλικού
 - Δεν διαθέτει επαρκείς μεθόδους διαχείρισης μεταδεδομένων
 - Δεν υποστηρίζει OAI-PMH αλλά το παλαιότερο Z39.50
 - Ελάχιστες δυνατότητες παραμετροποίησης
 - Ελλιπής προγραμματιστική διεπαφή



DSpace

- Υποστηρίζει αρχιτεκτονική τριών επιπέδων και παρέχει καλή υποστήριξη για την ανάπτυξη μιας ψηφιακής βιβλιοθήκης από ανεξάρτητα έγγραφα
- Κάνει χρήση σχεσιακής βάσης για αποθήκευση μεταδεδομένων ενώ το ψηφιακό υλικό αποθηκεύεται στο σύστημα αρχείων
- Υποστηρίζει DC και OAI αλλά όχι σε λεπτομερές επίπεδο
- Έχει περιορισμούς στη χρήση συσχετίσεων
- Θέτει ανυπέρβλητα προβλήματα για την υποστήριξη των συλλογών του ΕΚΠΑ
 - Δεν μπορεί εύκολα να επεκταθεί με σκοπό την παραμετροποίηση και προσαρμογή του
 - Πιθανόν να απαιτεί ακόμα και μεταβολή στον πηγαίο του κώδικα
 - Τίθεται θέμα συμβατότητας με επόμενες εκδόσεις



Fedora

- Βασίζεται στην «**F**lexible **E**xtensible **D**igital **O**bject **R**epository **A**rchitecture»
- Διατίθεται υπό την Mozilla Public License
- Βασισμένο σε Java, XML και Web Services
- Βασική οντότητα το ψηφιακό αντικείμενο που αποθηκεύεται σαν αρχείο XML (εμπεριέχει μοναδικό ID, μεταδεδομένα, datastreams και disseminators)
- Διαθέτει ανοικτές προγραμματιστικές διεπαφές (APIs) που αντιμετωπίζονται σαν Web Services
 - Διεπαφή διαχείρισης (API-M) και διεπαφή πρόσβασης (API-A)



Γιατί το Fedora

- Επικεντρωμένο στην παροχή επεκτάσιμης & ευέλικτης αρχιτεκτονικής
 - Ταιριάζει στους στόχους του ΕΚΠΑ
 - Έχει μεγάλες δυνατότητες επέκτασης
 - Υποστηρίζει την ανάπτυξη εξωτερικών modules που επικοινωνούν εύκολα με το σύστημα μέσω των προσφερόμενων διεπαφών (APIs)
 - Το μοντέλο ψηφιακών αντικειμένων του προσφέρει ενιαίο και εύχρηστο τρόπο αντιμετώπισης για όλες τις προδιαγεγραμμένες οντότητες
 - Δεν απαιτεί μεταβολές που να θέτουν προβλήματα συμβατότητας με επόμενες εκδόσεις
 - **Το Fedora χρησιμοποιείται ως πυρήνας του ΣΨΒ με όλη την απαιτούμενη λειτουργικότητα να υλοποιείται μέσω εξωτερικών modules**



Επεκτάσεις στο Fedora

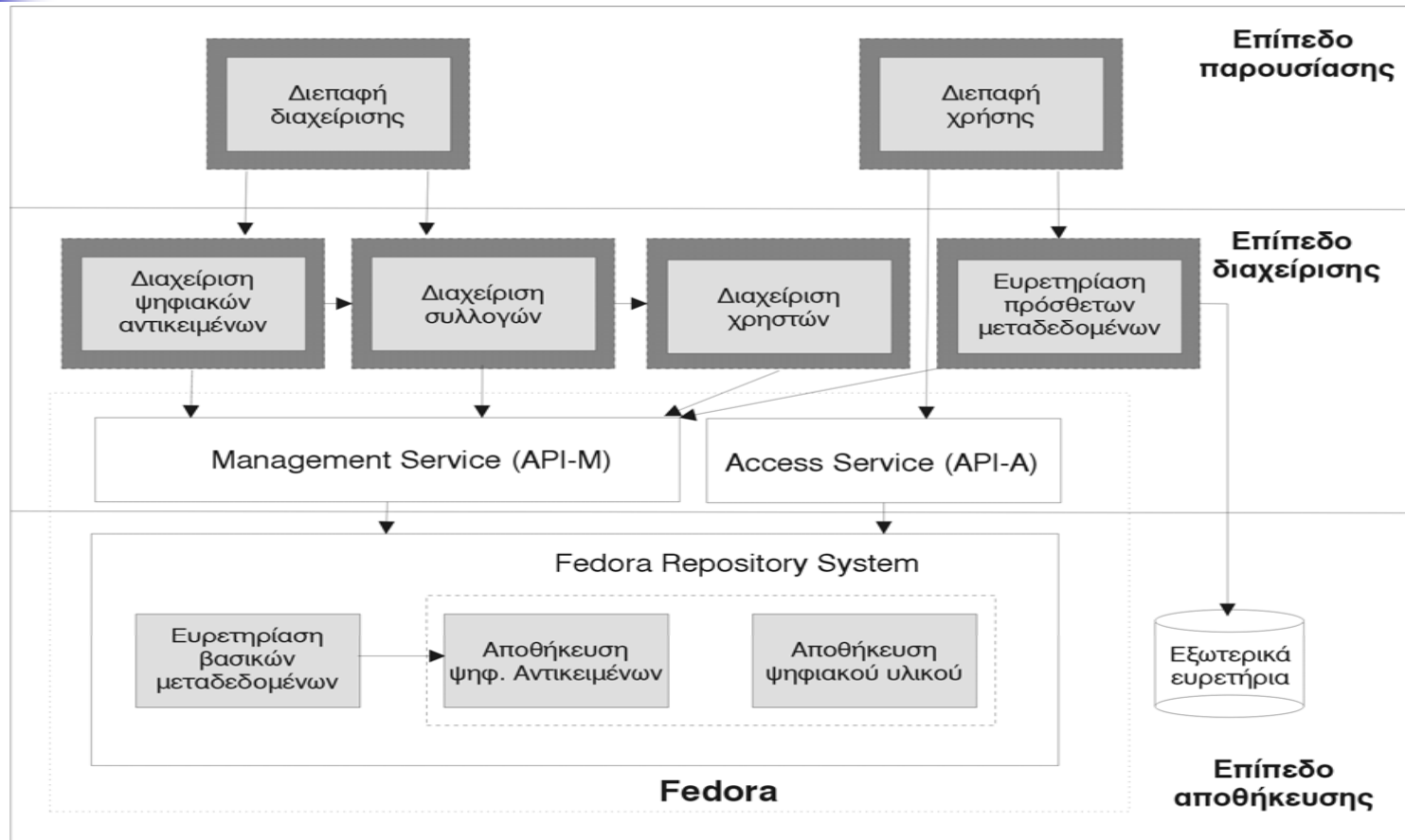
- Κύριος στόχος ο εμπλουτισμός της παρουσίασης των ψηφιακών αντικειμένων κάθε συλλογής:
 - Δομικά Στοιχεία (π.χ. κεφάλαιο, υποκεφάλαιο)
 - Συσχετίσεις (σχέσεις λογικής δομής και πλοήγησης, ακόμα και μεταξύ αντικειμένων διαφορετικών συλλογών)
 - Ευρετήρια όρων (βασισμένα στα μεταδεδομένα)
 - Ενιαία αναζήτηση ελευθέρου κειμένου βασισμένη σε metadata crosswalks
 - Λεπτομερής αναζήτηση (field search) στο πλαίσιο κάθε ξεχωριστής συλλογής
 - Η ευρετηρίαση για μεταδεδομένα πέραν του DC, υλοποιείται μέσω εξωτερικών υπηρεσιών που θα τις προσφέρουν οι εφαρμογές ανοικτού κώδικα Jakarta Lucene και Apache Xindice



Φιλοσοφία Σχεδιασμού

- Θέλουμε όλη την πληροφορία να αναπαρίσταται σαν ψηφιακό αντικείμενο του Fedora
- Επεκτείνουμε το μοντέλο αντικειμένων του Fedora και αναπαριστούμε τις νέες οντότητες που εισάγουμε ως «κανονικά» ψηφιακά αντικείμενα του Fedora
- Εισάγουμε την έννοια της συσχέτισης μεταξύ αντικειμένων ως κύριο τρόπο έκφρασης συμπεριφοράς και αναπαράστασης «σύνθετων αντικειμένων»
- Ενιαία αναπαράσταση και αποθήκευση όλων των αντικειμένων, επαναχρησιμοποίηση υπαρχόντων υποσυστημάτων του Fedora για την διαχείρισή τους

Προτεινόμενη Αρχιτεκτονική





Λεπτομέρειες αρχιτεκτονικής

- Το Fedora παρέχει υπηρεσίες στο επίπεδο αποθήκευσης και σε ποσοστό του επιπέδου διαχείρισης
- 4 βασικές υπηρεσίες υλοποιούνται στο επίπεδο διαχείρισης
 - Διαχείριση συλλογών: παραμετρική, βάση προτύπων (templates) που ορίζονται σε XML ώστε να υποστηρίζει δημιουργία συλλογών και υποσυλλογών, προσθήκη (χαρακτηρισμός) και ανάκτηση μεταδεδομένων, κλπ.
 - Διαχείριση ψηφιακών αντικειμένων (απλών και σύνθετων) και συσχετίσεων
 - Διαχείριση χρηστών (δημιουργία και επεξεργασία ομάδων χρηστών και καθορισμού δικαιωμάτων ανά συλλογή)
 - Ευρετηρίαση
- Στο επίπεδο παρουσίασης υλοποιούνται:
 - Η διεπαφή διαχείρισης (εφαρμογές διαχείρισης συλλογών, ψηφιοποίησης και καταλογογράφησης)
 - Η διεπαφή χρήσης που παρέχει πρόσβαση μέσω του παγκόσμιου ιστού και παρέχει τη δυνατότητα αναζήτησης και πλοήγησης



Τρέχουσα κατάσταση

- Υλοποίηση διαφόρων εξωτερικών modules
 - Διαχείριση αντικειμένων και διαχείριση συλλογών
- Ανάπτυξη των υπηρεσιών καταλογογράφησης και διαχείρισης και των αντίστοιχων ροών εργασίας
- Επικέντρωση αρχικά στην παραμετροποίηση για τρεις ψηφιακές συλλογές
 - Λαογραφίας
 - Βυζαντινών κωδίκων
 - Ιστορικού Αρχείου



Σχέδια για το άμεσο μέλλον

- Υλοποίηση των επεκτάσεων του υποσυστήματος παρουσίασης για τους τελικούς χρήστες
- Ενεργοποίηση της ευρετηρίασης πρόσθετων μεταδεδομένων
- Σταδιακή ενσωμάτωση και των υπολοίπων συλλογών
- Το ΥΚΒ θα προχωρεί σε βελτίωση της λειτουργικότητας με βάση τις εκάστοτε ανάγκες του ΕΚΠΑ