

Ορισμός, Προδιαγραφές και Λειτουργικό Μοντέλο για ένα γενικού-σκοπού
Ψηφιακό Απόθεμα

Δημήτριος Α. Κουτσομητρόπουλος (kotsomit@hpclab.ceid.upatras.gr)
Εργαστήριο Πληροφοριακών Συστημάτων Υψηλών Επιδόσεων
Οκτώβριος 2002

1. Ορίζοντας το Ψηφιακό Απόθεμα

Ο αυστηρός ορισμός της έννοιας «ψηφιακό απόθεμα» είναι ένα θέμα στο οποίο χωρά πολλή συζήτηση και είναι σχεδόν σίγουρο ότι δεν πρόκειται να επιτευχθεί ομοφωνία για έναν κοινό ορισμό. Μια από τις βασικές αιτίες της αμφισημίας αυτής είναι ότι ένα ψηφιακό απόθεμα χαρακτηρίζεται κατά μείζονα λόγο από το είδος του περιεχομένου που καθιστά διαθέσιμο και από τους λόγους που υπαγορεύουν τη δημιουργία του και τη λειτουργικότητά του. Ωστόσο, δεν απαιτούνται εξειδικευμένες γνώσεις για να μπορέσει κάποιος να κατανοήσει την ορθότητα του ακόλουθου διαισθητικού ορισμού:

Ορισμός 1: Ένα ψηφιακό απόθεμα είναι μια συλλογή από ψηφιακές οντότητες που υπόκειται στις ακόλουθες τρεις λειτουργίες: Εισαγωγή, διαγραφή και ανάκτηση.

Ο παραπάνω ορισμός είναι αρκετά αφαιρετικός όσο και απλός και είναι προφανές ότι παρέχει το ελάχιστο σύνολο απαιτήσεων για οποιοδήποτε ψηφιακό απόθεμα. Ένας κάπως πιο εκλεπτυσμένος ορισμός είναι ο ακόλουθος:

Ορισμός 2 (RLG) : Ένα ψηφιακό απόθεμα είναι μια οργάνωση που έχει την ευθύνη της μακροπρόθεσμης διατήρησης ψηφιακών πόρων, καθώς και της διάθεσής τους σε συγκεκριμένες κοινότητες που έχουν συνάψει συμφωνία με τον Παραγωγό (του περιεχομένου) και την αρχή διαχείρισης του αποθέματος.

Ο ορισμός αυτός εισάγει έννοιες, όπως «μακροπρόθεσμη διατήρηση», που δεν είναι διαισθητικά και άμεσα αντιληπτές. Ουσιαστικά προέρχεται από τον ορισμό του «Ανοιχτού Αρχειακού Πληροφοριακού Συστήματος» (Open Archival Information System – OAIS). Η κατανόηση του ορισμού αυτού όμως απαιτεί πρώτα την επεξήγηση της ιδιαίτερης ορολογίας που χρησιμοποιείται στο OAIS.

Ας δούμε λοιπόν το ψηφιακό απόθεμα ως ένα κλειστό σύστημα. Τότε το ψηφιακό απόθεμα αλληλεπιδρά με τρεις οντότητες του εξωτερικού περιβάλλοντος (βλ. Σχ.1): τον Παραγωγό, τον Καταναλωτή και τη Διαχείριση.



Σχήμα 1: Μοντέλο Αλληλεπίδρασης Ψηφιακού Αποθέματος με το Περιβάλλον

Παραγωγός: Είναι ο ρόλος που αντιστοιχεί σε εκείνα τα πρόσωπα ή client συστήματα, που παρέχουν την πληροφορία που πρόκειται να διατηρηθεί.

Διαχείριση: Είναι ο ρόλος που αντιστοιχεί σε εκείνους που θέτουν την γενική πολιτική του αποθέματος σαν ένα μόνο δομικό στοιχείο μιας ευρύτερης πολιτικής. Με άλλα λόγια ο διαχειριστικός έλεγχος του αποθέματος αποτελεί μία μόνο από τις ευθύνες της Διαχείρισης. Η Διαχείριση δεν ασχολείται με τις καθημερινές αρχειακές λειτουργίες. Η καθημερινή διαχείριση του αποθέματος ορίζεται εσωτερικά στο σύστημα.

Καταναλωτής: Είναι ο ρόλος που αντιστοιχεί σε εκείνα τα πρόσωπα ή client συστήματα, που αλληλεπιδρούν με τις υπηρεσίες του αποθέματος για να βρουν και να ανακτήσουν τη διατηρούμενη πληροφορία που τους ενδιαφέρει. Μια ειδική κλάση Καταναλωτών είναι η Προσδιορισμένη Κοινότητα. Η Προσδιορισμένη Κοινότητα είναι το σύνολο των Καταναλωτών που θα πρέπει να μπορούν να κατανοούν την πληροφορία που διατηρείται.

Μένει τώρα να εξηγήσουμε τη εννοούμε με τον όρο «Μακροπρόθεσμη Διατήρηση»:

Μακροπρόθεσμα: Μια χρονική περίοδος αρκετά μεγάλη ώστε να δημιουργεί ανησυχία για τις επιπτώσεις των τεχνολογικών αλλαγών, όπως υποστήριξη για νέα μέσα και τύπους δεδομένων και των αλλαγών στην κοινότητα χρηστών, πάνω στις πληροφορίες ενός αποθέματος. Η περίοδος αυτή εκτείνεται απροσδιόριστα στο μέλλον.

Μακροπρόθεσμη Διατήρηση: Η ενέργεια της διατήρησης της πληροφορίας, με ορθή και Ανεξάρτητα Κατανοήσιμη μορφή, μακροπρόθεσμα.

Ανεξάρτητα Κατανοήσιμος: Χαρακτηριστικό της πληροφορίας εκείνης που έχει επαρκή τεκμηρίωση, ώστε να επιτρέπει την κατανόηση και χρήση της από την Προσδιορισμένη Κοινότητα, χωρίς τη βοήθεια ειδικών πόρων που δεν είναι ευρέως διαθέσιμοι.

Η ανάγκη της μακροπρόθεσμης διατήρησης έχει από καιρό αναγνωριστεί από σημαντικούς φορείς υλοποίησης ψηφιακών αποθεμάτων (βιβλιοθήκες, οργανισμούς ψηφιοποίησης, consortia προτυποποίησης κ.α.) καθώς και από την Ευρωπαϊκή Ένωση στο θέμα της ψηφιοποίησης της πολιτιστικής κληρονομιάς (βλ. Ισπανική Προεδρία). Η περαιτέρω ανάλυση της σπουδαιότητας της μακροπρόθεσμης διατήρησης είναι έξω από τους σκοπούς της παρούσας αναφοράς.

Είμαστε πλέον έτοιμοι να δώσουμε τον ακόλουθο ορισμό του Ανοιχτού Αρχειακού Πληροφοριακού Συστήματος:

Ορισμός 3 (OAIS): Ένα ψηφιακό απόθεμα είναι ένα αρχείο που στοχεύει στη διατήρηση της ψηφιακής πληροφορίας για προσπέλαση και χρήση από μια Προσδιορισμένη Κοινότητα και ικανοποιεί τις ακόλουθες απαιτήσεις:

- Διαπραγματεύεται και αποδέχεται τις κατάλληλες πληροφορίες από τους Παραγωγούς πληροφορίας.
- Αποκτά επαρκή έλεγχο στην παρεχόμενη πληροφορία σε τέτοιο βαθμό, ώστε να διασφαλίζεται η Μακροπρόθεσμη Διατήρησή της.

- *Καθορίζει, είτε αυτόνομα είτε σε συνεργασία με άλλα μέρη, ποιες κοινότητες θα αποτελούν την Προσδιορισμένη Κοινότητα και επομένως θα πρέπει να μπορεί να κατανοεί την παρεχόμενη πληροφορία.*
- *Διασφαλίζει ότι η πληροφορία που διατηρεί είναι Ανεξάρτητα Κατανοήσιμη από την Προσδιορισμένη Κοινότητα. Με άλλα λόγια η Προσδιορισμένη Κοινότητα θα πρέπει να μπορεί να κατανοεί την πληροφορία, χωρίς να χρειάζεται την βοήθεια των ειδικών που την παρήγαγαν.*
- *Ακολουθεί τεκμηριωμένες πολιτικές και διαδικασίες που εξασφαλίζουν ότι η πληροφορία θα διατηρείται ενάντια σε κάθε λογικό απρόοπτο και θα διαμοιράζεται σε πιστοποιημένα αντίγραφα του πρωτοτύπου ή όσο το δυνατόν πιο ανιχνεύσιμα προς το πρωτότυπο.*
- *Καθιστά την πληροφορία που διατηρείται διαθέσιμη στην Προσδιορισμένη Κοινότητα.*

Είναι σαφές ότι καθένας από τους παραπάνω ορισμούς θέτει διαφορετικές προδιαγραφές για το ψηφιακό απόθεμα. Μάλιστα οι παραπάνω ορισμοί, ενώ παραμένουν αρκετά γενικοί και αόριστοι όσον αφορά τον τύπο της πληροφορίας και την προοριζόμενη χρήση της, ακολουθούν μια εξελικτική πορεία όσον αφορά την μεθοδολογία διατήρησης της πληροφορίας στο απόθεμα και τους τρόπους αλληλεπίδρασης του τελευταίου με το περιβάλλον. Για παράδειγμα, ο Ορισμός 1 καλύπτεται ακόμα και από μια αδόμητη συλλογή δεδομένων (δηλ. όχι αναγκαστικά Βάση Δεδομένων), συνοδευόμενη από μια διεπαφή εισαγωγής / διαγραφής και μια διεπαφή αναζήτησης / ανάκτησης. Ο Ορισμός 2 εισάγει την έννοια της μακροπρόθεσμης διατήρησης, ενώ τέλος ο Ορισμός 3 εκλεπτύνει ακόμα περισσότερο θίγοντας έμμεσα την ποιότητα των προσφερόμενων από το απόθεμα υπηρεσιών και το θέμα των πνευματικών δικαιωμάτων.

Φυσικά, οι παραπάνω ορισμοί δεν εξαντλούν όσα μπορεί κανείς να εννοήσει με τον όρο «ψηφιακό απόθεμα» και που προκύπτουν από τις συγκεκριμένες αναγκαιότητες και προτεραιότητές του. Επιτυγχάνουν όμως να θίξουν μια σειρά ζητημάτων που σχετίζονται με τη διατήρηση και τη διάθεση του περιεχομένου, που θα πρέπει να ληφθούν σοβαρά υπόψη από οποιονδήποτε πρόκειται να υλοποιήσει ένα ψηφιακό απόθεμα. Αξίζει τέλος να σημειωθεί ότι κανένας από τους ορισμούς που δόθηκαν δεν υπαγορεύει συγκεκριμένο σχεδιασμό ή υλοποίηση ενός ψηφιακού αποθέματος.

2. Προδιαγραφές για ένα Ψηφιακό Απόθεμα

Από την προηγούμενη συζήτηση περί του ορισμού του ψηφιακού αποθέματος, αλλά και από τη μελέτη της τρέχουσας ερευνητικής δραστηριότητας στο χώρο των ψηφιακών αποθεμάτων και ψηφιακών βιβλιοθηκών, προκύπτουν οι παρακάτω άξονες προδιαγραφών για ένα ψηφιακό απόθεμα:

- **Μακροπρόθεσμη διατήρηση / προσπέλαση στο περιεχόμενο του αποθέματος:** Μπορεί να επιτευχθεί αναπτύσσοντας ένα απόθεμα συμβατό με τις προδιαγραφές του OAIS. Το περιεχόμενο θα πρέπει να χαρακτηρίζεται από

μόνιμη δυνατότητα προσπέλασης (π.χ. μέσω persistent URIs) και το απόθεμα θα πρέπει να μπορεί να διατηρεί πολλαπλές εκδόσεις του περιεχομένου, όπως αυτό τροποποιείται / ενημερώνεται στην πορεία του χρόνου (persistence).

- **Μεταδεδομένα:** Θα πρέπει να προβλέπονται μεταδεδομένα για όλο τον κύκλο ζωής του ψηφιακού περιεχομένου. Σκοπός είναι η περιγραφή του περιεχομένου των ψηφιακών πόρων του αποθέματος, η υποβοήθηση της διαχείρισής του και η διευκόλυνση της προσπέλασής του, ακόμα και μακροπρόθεσμα (descriptive, administrative, preservation metadata). Για να έχουν νόημα, τα μεταδεδομένα θα πρέπει να ακολουθούν κάποιο ευρέως χρησιμοποιούμενο πρότυπο (προτείνεται Dublin Core στη γενική περίπτωση, MPEG-7 για πολυμεσικό περιεχόμενο, DIG-35 για ψηφιακές εικόνες και METS (Metadata Encoding and Transmission Standard - για την κωδικοποίηση και διαχείριση των παραπάνω). Επίσης, σε ορισμένες περιπτώσεις, τα μεταδεδομένα δεν θα πρέπει να περιορίζονται μόνο στις αντικείμενα του αποθέματος, αλλά θα πρέπει να υποστηρίζουν και ένα ανώτερο επίπεδο αφαίρεσης, το επίπεδο συλλογών (collection level, βλ. προδιαγραφές Dublin Core – CLD).
- **Διαλειτουργικότητα / Δυνατότητα εισόδου-εξόδου:** Η διαλειτουργικότητα του συστήματος του ψηφιακού αποθέματος θα επιτευχθεί ακολουθώντας γνωστά πρότυπα κατά την ανάπτυξη της λειτουργικότητας του αποθέματος. Ένα απ' αυτά είναι η ανεξάρτητη πλατφόρμας γλώσσα XML (& XML Schema), ενώ ισχυρά προτείνεται η υλοποίηση του πρωτοκόλλου OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) για τη δυνατότητα μαζικής εισαγωγής / εξαγωγής των μεταδεδομένων του αποθέματος. Η υποστήριξη επίσης του πρωτοκόλλου Z39.50 θεωρείται κρίσιμης σημασίας, όταν πρόκειται για την διαφανή και απομεμακρυσμένη αναζήτηση σε μεγάλο όγκο εγγραφών του αποθέματος. Η διαλειτουργικότητα και προσπελασιμότητα του ψηφιακού αποθέματος ενισχύεται επίσης, εκθέτοντας τις υπηρεσίες του ως Υπηρεσίες Ιστού, πράγμα που πρακτικά σημαίνει περιγραφή τους μέσω της WSDL γλώσσας και δήλωσής τους σε κάποιο UDDI registry. Το σημαντικό με το UDDI είναι ότι επιτρέπει την αυτοματοποιημένη ανεύρεση (και ενδεχομένως αξιοποίηση) μιας Υπηρεσίας Ιστού από τη μηχανή, κατά τον ίδιο τρόπο που οι φυσικοί χρήστες χρησιμοποιούν σήμερα τις μηχανές αναζήτησης.
- **Ασφάλεια / Πιστοποίηση Χρηστών:** Είναι ευνόητο ότι το ψηφιακό απόθεμα δεν θα είναι προσπελάσιμο από οποιαδήποτε κοινότητα Καταναλωτών, καθώς στοχεύει στην Προσδιορισμένη Κοινότητα. Πρακτικός τρόπος για να επιτευχθεί αυτό είναι η θέσπιση πολιτικών προσπέλασης για κάθε Καταναλωτή / Κοινότητα Καταναλωτών, η πιστοποίηση τους μέσω ζευγών login – password ή / και ψηφιακών πιστοποιητικών και η κρυπτογραφημένη πρόσβαση στις υπηρεσίες του αποθέματος (SSL,...).
- **Διαχείριση / Αναπαράσταση Γνώσης:** Το περιεχόμενο του αποθέματος δεν θα περιορίζεται αποκλειστικά σε ένα θεματικό χώρο (domain), αλλά μπορεί να ανήκει σε πολλούς χώρους ή συνδυασμούς τους. Είναι επομένως σκόπιμο, το περιεχόμενο να μην είναι αδόμητο, αλλά ιεραρχημένο σημασιολογικά με σαφή και δομημένο τρόπο. Με άλλα λόγια προτείνεται η κατάρτιση

οντολογιών για κάθε θεματικό χώρο του περιεχόμενου. Για παράδειγμα, στο χώρο της πολιτιστικής κληρονομιάς, προτείνεται η αξιοποίηση της προτεινόμενης από το CIDOC οντολογίας (CRM-Conceptual Reference Model). Η βασικότερη λειτουργικότητα που παρέχουν οι οντολογίες είναι ότι επιτρέπουν την μετατόπιση της αναζήτησης / ανάκτησης από το επίπεδο των λέξεων-κλειδίων στο επίπεδο γνώσης. Έτσι το σύστημα μπορεί να υποβοηθήσει το χρήστη στην αναζήτηση του, υποστηρίζοντας αυτόματη εξαγωγή συμπερασμάτων (automated reasoning) ακόμα κι όταν η σχετική πληροφορία δεν ορίζεται ρητά στα μεταδεδομένα του περιεχομένου του αποθέματος. Σε ένα δεύτερο επίπεδο προβλέπεται η ανάπτυξη Έξυπνων Πρακτόρων (Intelligent Agents) οι οποίοι επικοινωνούν μεταξύ τους, ανταλλάσσουν πληροφορίες, υποβάλλουν έξυπνα ερωτήματα στο απόθεμα και εξελίσσουν την οντολογία του. Οντολογίες μπορούν επίσης να εφαρμοστούν και στις διαχειριστικές πτυχές του αποθέματος, όπως η ABC για την οργάνωση των συμβάντων που λαμβάνουν χώρα στο απόθεμα, στην πορεία του χρόνου. Για την ανάπτυξη των οντολογιών παραδοσιακά χρησιμοποιείται το RDF, ωστόσο προτείνεται η DAML+OIL και η ακόμα πιο πρόσφατη OWL, γλώσσες που σχεδιάστηκαν ειδικά για οντολογίες.

- **Προστασία Πνευματικών Δικαιωμάτων:** Η εξασφάλιση των δικαιωμάτων των δημιουργών και η οικονομική αξιοποίηση του αποθέματος καθιστούν επιτακτική την ανάγκη της διαχείρισης και της κωδικοποίησης των πνευματικών δικαιωμάτων στο περιεχόμενο. Η μέθοδος της υδατογράφησης όχι μόνο για ψηφιακές εικόνες, αλλά και για κάθε τύπο πολυμεσικού αντικειμένου είναι ευρέως χρησιμοποιούμενη. Σε επίπεδο μεταδεδομένων αναφέρεται ενδεικτικά η XML-based Γλώσσα Περιγραφής Πνευματικών Δικαιωμάτων του MPEG-21 καθώς και το πακέτο διασφάλισης XML του W3C (XML Encryption, XML Key Management, XML Signature).

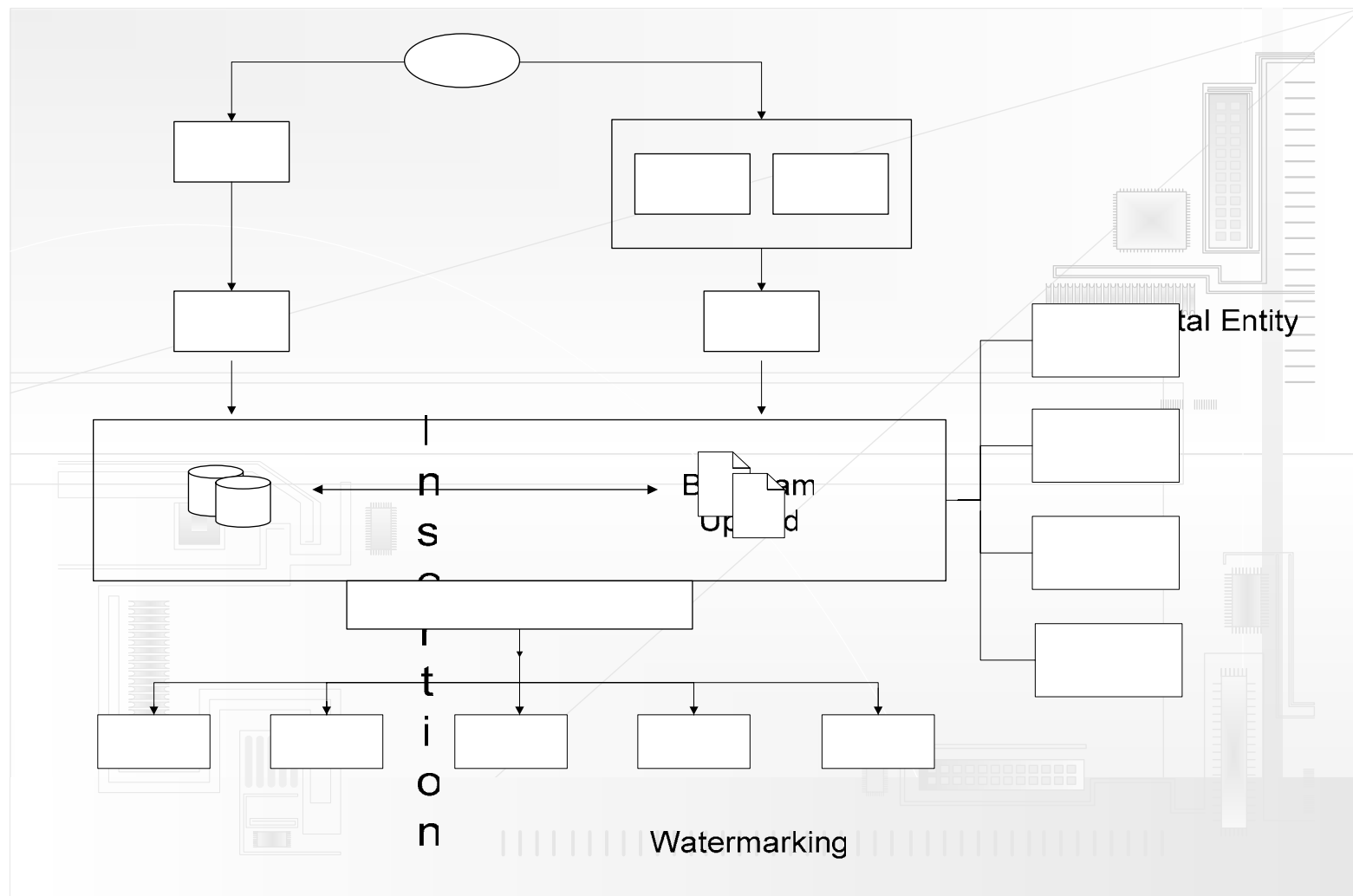
3. Λειτουργικό Μοντέλο Ψηφιακού Αποθέματος

Εφαρμόζοντας τις παραπάνω προδιαγραφές, προκύπτει το ακόλουθο *Λειτουργικό Μοντέλο (Functional Model)* του ψηφιακού αποθέματος (βλ. Σχ.2). Το λειτουργικό μοντέλο αποτελεί ουσιαστικά εκλέπτυνση του ορισμού του ψηφιακού αποθέματος ως κλειστού συστήματος, περιγράφοντας αναλυτικά τους τρόπους εσωτερικής οργάνωσής του, καθώς και τις μεθόδους αλληλεπίδρασης με τις εξωτερικές οντότητες Παραγωγό, Καταναλωτή και Διαχείριση. Αξίζει να σημειωθεί ότι το λειτουργικό μοντέλο που προτείνεται εδώ δεν είναι απαραίτητα συμμορφωμένο με το OAIS μοντέλο αναφοράς. Αντίθετα η συμμόρφωση με τις προδιαγραφές του OAIS είναι προαιρετική, μαζί με άλλες σχεδιαστικές επιλογές (βλ. Παρ. 3.3).

Το λειτουργικό μοντέλο χωρίζεται σε 4 επίπεδα: Την Εισαγωγή, το Απόθεμα, τη Διαχείριση και την Κατανάλωση.

3.1. Εισαγωγή

Ο Παραγωγός οφείλει να παρέχει την οντότητα προς διατήρηση σε ψηφιακή μορφή. Είναι επίσης ευθύνη του Παραγωγού η παροχή ψηφιακών οντοτήτων σε αποδεκτές από το απόθεμα, προσπελάσιμες και απαλλαγμένες από λάθη μορφές (*Submission Agreement*, βλ. OAIS). Από τη στιγμή που ένα ψηφιακό αντικείμενο είναι διαθέσιμο, εισάγεται στο απόθεμα σε δύο παράλληλες συνιστώσες:



Σχήμα 2: Λειτουργικό Μοντέλο Ψηφιακού Αποθέματος

Repository (Actual Storage)

- *Δυναδική Μεταφόρτωση*: Η ψηφιακή (δυναδική) αναπαράσταση του αντικειμένου αποθηκεύεται στο απόθεμα. Πρόκειται ουσιαστικά για την παροχή μιας διεπαφής με την οποία ο Παραγωγός θα μπορεί να μεταφορτώνει αρχεία στο απόθεμα. Το δυναδικό αρχείο, αφού υδατογραφηθεί κατάλληλα, αποθηκεύεται στην ΒΔ του αποθέματος.
- *Προσθήκη Μεταδεδομένων*: Ανάλογα με τον τύπο του περιεχομένου, έχουν προαποφασιστεί τα σχήματα μεταδεδομένων που θα ακολουθηθούν. Με βάση αυτά, θα παρέχεται στον Παραγωγό μια σειρά από φόρμες με τα απαιτούμενα πεδία μεταδεδομένων προς συμπλήρωση (*Χειρωνακτική Εισαγωγή*), ενώ θα καταβάλλεται προσπάθεια για όσο το δυνατόν πιο επαυξημένη αυτόματη εξαγωγή μεταδεδομένων (*Αυτόματη Εξαγωγή*), όπου αυτό είναι δυνατό (π.χ. τα χρώματα σε μια εικόνα). Στη φάση αυτή εισάγονται επίσης και τα μεταδεδομένα που αφορούν στα πνευματικά δικαιώματα του περιεχομένου. Πρώτου τα μεταδεδομένα αποθηκευτούν στο απόθεμα, κωδικοποιούνται με κατάλληλο τρόπο και, ενδεχομένως, κρυπτογραφούνται. Έτσι τα μεταδεδομένα μπορούν: α) να οργανωθούν σε διαφορετικά XML αρχεία ανάλογα με το σχήμα που ακολουθούν β) να κωδικοποιηθούν με βάση το METS πρότυπο γ) να αποθηκευτούν σχεσιακά σε πεδία της ΒΔ. Να σημειωθεί ότι η τελευταία επιλογή καταργεί τη δυνατότητα διαχωρισμού του περιεχομένου από την περιγραφή του (βλ. και *Απόθεμα*).

3.2. Απόθεμα

Πρόκειται για το επίπεδο όπου λαμβάνει χώρα η πραγματική αποθήκευση του περιεχομένου. Το απόθεμα δέχεται ως είσοδο τα παράγωγα της Εισαγωγής, τροφοδοτεί την Κατανάλωση και επηρεάζεται εξωτερικά από τις επιλογές της Διαχείρισης.

Καθώς η εισαγωγή μιας ψηφιακής οντότητας γίνεται σε δύο συνιστώσες (αρχείο + μεταδεδομένα), η αποθήκευση της οντότητας στο απόθεμα ακολουθεί παρόμοια λογική. Θα πρέπει λοιπόν να αντιμετωπιστεί το πρόβλημα της αποθήκευσης δύο χωριστών εννοιολογικά συνιστωσών: του δυναδικού αρχείου, όπου βρίσκεται δυναδικά κωδικοποιημένη η ψηφιακή οντότητα, και των συνοδευτικών μεταδεδομένων της. Ανάλογα με το κατά πόσο οι συνιστώσες αυτές είναι και φυσικά διαχωρισμένες, προκύπτουν οι ακόλουθες σχεδιαστικές επιλογές:

- *Relationally-biased λύση*: Αποθήκευση τόσο των αρχείων, όσο και των μεταδεδομένων σε πολυμεσική σχεσιακή Βάση Δεδομένων. Τα αρχεία αποθηκεύονται σαν BLOBS και σχεδιάζεται διάγραμμα Οντοτήτων-Συσχετίσεων (ER), που θα οδηγήσει στην κατάρτιση των αντίστοιχων πινάκων στη ΒΔ. Τα μεταδεδομένα είναι αποθηκευμένα σε πεδία των παραπάνω πινάκων. Το πρόβλημα της διάχυσης XML πληροφορίας που προκύπτει άμεσα, αφού καταργούνται τα XML έγγραφα, αντιμετωπίζεται με τη δυναμική κατ' απαίτηση (on-demand) παραγωγή τους από τα πεδία της ΒΔ. Η λύση αυτή έχει το πλεονέκτημα ότι τα μεταδεδομένα βρίσκονται άμεσα συνδεδεμένα με το περιεχόμενο που περιγράφουν, μέσω των σχέσεων του ER μοντέλου.

- *Document-biased λύση*: Στη λύση αυτή, τα μεταδεδομένα χωρίζονται και φυσικά από το περιεχόμενο. Τα ψηφιακά αρχεία αποθηκεύονται στην πολυμεσική ΒΔ, μαζί ενδεχομένως με κάποιες άλλες πληροφορίες (όπως labels, δείκτες, identifiers κλπ), ενώ τα μεταδεδομένα αποθηκεύονται χωριστά σε ένα XML Document Repository. Το τελευταίο μπορεί είτε να αναπτυχθεί ξεχωριστά, είτε να βασιστεί σε κάποιο εμπορικό native XML DBMS (π.χ. Tamino). Η ανάγκη της σύνδεσης μεταξύ των αντικειμένων που είναι αποθηκευμένα στη ΒΔ και των μεταδεδομένων τους, οδηγεί σε επιβάρυνση της επιχειρησιακής λογικής (business logic), καθώς δεν υπάρχει άλλος τρόπος συσχέτισης μεταξύ των συνιστωσών. Με άλλα λόγια, απαιτείται η συγγραφή επιπλέον κώδικα και η χρήση μοναδικών αναγνωριστικών τόσο στη ΒΔ, όσο και στα XML έγγραφα για τη διασφάλιση της συνέπειας (consistency), ότι δηλαδή τα XML έγγραφα δείχνουν στα αντικείμενα εκείνα που πραγματικά περιγράφουν. Η αποδοτική ανάκτηση πληροφοριών από μια συλλογή XML εγγράφων αντιμετωπίζεται με τις αντίστοιχες προς την SQL XML γλώσσες έρωτησης: XQuery και XQL.
- *Unbiased λύση*: Τέλος, μια μέση λύση είναι η χρήση μιας σχεσιακής ΒΔ με εγγενή XML υποστήριξη (Μέχρι στιγμής η Oracle 9i Release 2 και, λιγότερο, η IBM DB2 με XML Extenders). Έτσι, η σύνδεση περιεχομένου – μεταδεδομένων είναι πιο άμεση απ’ ότι στην προηγούμενη λύση, αφού ουσιαστικά διαχειρίζονται από το ίδιο DMBS. Την ίδια στιγμή, τα μεταδεδομένα παραμένουν ικανοποιητικά διαχωρισμένα από το περιεχόμενο, καθώς αποθηκεύονται ανεξάρτητα σε XML έγγραφα και ακολουθούν το δικό τους σχήμα.

Σε κάθε περίπτωση, τα ψηφιακά αντικείμενα συνδέονται με τα XML έγγραφα με σχέση 1:n. Έτσι, ένα XML αρχείο μπορεί να αντιστοιχεί σε πολλά ψηφιακά αντικείμενα, καλύπτοντας την ανάγκη για περιγραφή του περιεχομένου και σε επίπεδο συλλογών.

Η κατάλληλη σχεδιαστική επιλογή για τον τρόπο αποθήκευσης περιεχομένου-μεταδεδομένων κρίνεται ιδιαίτερα κρίσιμη, καθώς καλείται να εξασφαλίσει την αποδοτική προσπέλαση του αποθέματος, την εξέλιξή του και την διατήρησή του στην πορεία του χρόνου.

Καταρχάς, η χρήση ενός σχεσιακού DMBS κρίνεται απαραίτητη, τουλάχιστον για τα ψηφιακά αρχεία, καθώς παρέχει τη συνηθέστερη και πιο αποδοτική λύση στο πρόβλημα της μαζικής, αξιόπιστης και εύρωστης αποθήκευσης και ανάκτησης πληροφορίας. Ακόμα, η ΒΔ χαρακτηρίζεται ως *πολυμεσική*, με την έννοια ότι θα πρέπει να μπορεί να υποστηρίξει την αποδοτική αποθήκευση και διαχείριση μεγάλων, πολυμεσικών αρχείων όλων των τύπων (ήχος, video, εικόνα...). Τα περισσότερα σύγχρονα DBMS μπορούν να θεωρηθούν πολυμεσικά με την έννοια αυτή. Σε ένα δεύτερο επίπεδο, μια πολυμεσική βάση δεδομένων θα πρέπει να υποστηρίζει προηγμένες αναζητήσεις με βάση το περιεχόμενο, όπως π.χ. αναζητήσεις με βάση το χρώμα, το σχήμα ή την υφή μιας ψηφιακής εικόνας. Παράδειγμα μιας τέτοιας ΒΔ είναι η IBM DB2 με Multimedia Extenders.

Ωστόσο, το πρόβλημα των προηγμένων αναζητήσεων σε πολυμεσικό περιεχόμενο μπορεί να αντιμετωπιστεί και με τη βοήθεια των μεταδεδομένων. Για παράδειγμα, το MPEG-7 προβλέπει μεταδεδομένα που περιγράφουν λεπτομερώς τα χαρακτηριστικά ενός πολυμεσικού αντικειμένου (σχήμα,

χρώμα...), τα οποία μάλιστα μπορούν να εξαχθούν αυτόματα. (βλ. *Προσθήκη Μεταδεδομένων – Αυτόματη Εξαγωγή*). Επιπλέον, τα μεταδεδομένα αυτά χρειάζεται να εξαχθούν μόνο μια φορά, κατά την φάση της Εισαγωγής του αντικειμένου στο Απόθεμα.

Ο διαχωρισμός του περιεχομένου από την περιγραφή του, ακόμα και φυσικά, φαίνεται να αποτελεί κυρίαρχη πρακτική σε συναφείς πανεπιστημιο-κεντρικές πρωτοβουλίες ανάπτυξης αποθεμάτων (π.χ. MIT-DSpace, Cornell-Fedora). Τα βασικότερα πλεονεκτήματα μιας τέτοιας προσέγγισης συνοψίζονται παρακάτω:

- *Αυτονομία των πληροφοριών του περιεχομένου*: Αφού η συνοδευτικές πληροφορίες ενός ψηφιακού αντικειμένου, δηλαδή τα μεταδεδομένα του, είναι διαχωρισμένα από το αντικείμενο καθαυτό, χαρακτηρίζονται από απόλυτη αυτονομία. Έτσι, ακόμα και σε περίπτωση αποτυχίας ή καταστροφής της ΒΔ, το ψηφιακό απόθεμα θα είναι σε θέση να διαχέει τις πληροφορίες του αντικειμένου, παρόλο που αυτό δεν θα είναι προσπελάσιμο. Ένα παράδειγμα είναι μια ιστοσελίδα όπου η φωτογραφία που ανακτά ο χρήστης δεν εμφανίζεται, αλλά όλες οι σχετικές μ' αυτήν πληροφορίες (τίτλος, όνομα δημιουργού, περιγραφή, ακόμα και ένα URL προς κάποιο αντίγραφο της εικόνας κάπου άλλου) είναι προσπελάσιμες.
- *Υποστήριξη XML-oriented προτύπων και Διαλειτουργικότητα*: Αρκετά πρότυπα κατανάλωσης, που μάλιστα τυγχάνουν ευρείας αποδοχής σήμερα (όπως π.χ. το OAI), απαιτούν την παροχή της πληροφορίας σε XML μορφή. Έτσι, η χωριστή αποθήκευση XML εγγράφων προκύπτει ως πιο φυσική και κυρίως, πιο αποδοτική λύση, ιδιαίτερα στην περίπτωση της μαζικής κατανάλωσης, έναντι της δυναμικής παραγωγής XML εγγράφων από μια ΒΔ. Ακόμα, η platform-independent φύση της XML, καθώς και το human / machine readable χαρακτηριστικό της, σε συνδυασμό με την ευρεία αποδοχή και αξιοποίησή της γλώσσας τα τελευταία χρόνια, την καθιστούν ιδανική για την επίτευξη διαλειτουργικότητας με οποιοδήποτε σχεδόν σύγχρονο πληροφοριακό σύστημα.
- *Ευκολότερη αλλαγή του σχήματος μεταδεδομένων*: Η XML Schema αποτελεί ένα ισχυρό εργαλείο για τον σχεδιασμό των σχημάτων των XML εγγράφων και τη θέσπιση των επιθυμητών περιορισμών. Με τη χρήση της XML Schema, η διαδικασία αλλαγής ή συγχώνευσης σχημάτων μεταδεδομένων απλοποιείται αισθητά σε σχέση με την τροποποίηση του σχήματος μιας ΒΔ. Επιπλέον αποτελεί τη μόνη αποδοτική λύση στην περίπτωση ενός εξελισσόμενου, ανοιχτού σχήματος (semi-structured), όπως συχνά συμβαίνει με τα σχήματα των μεταδεδομένων.

Από την άλλη, ένας πλήρης φυσικός διαχωρισμός περιεχομένου-μεταδεδομένων επισύρει, όπως αναφέρθηκε, την ευθύνη της διατήρησης της συνέπειας μεταξύ του περιεχομένου και της περιγραφής του και τη σχετική προγραμματιστική επιβάρυνση. Τέλος, είναι ακόμα ανοιχτό εάν μια σχεσιακή ΒΔ είναι αποδοτικότερη, συνολικά, από μια native XML ΒΔ (π.χ. οι ερωτήσεις απαντώνται ταχύτερα;).

3.3. Διαχείριση

Η Διαχείριση εδώ έχει την έννοια που δόθηκε στον ορισμό του O AIS, δηλαδή είναι εξωτερική προς το απόθεμα και δεν έχει να κάνει με την καθημερινή συντήρηση / διαχείριση του περιεχομένου του. Σύμφωνα με τον ορισμό του O AIS, η Διαχείριση τυπικά περιλαμβάνει τη διαπραγμάτευση των οικονομικών πόρων, τη διενέργεια τακτικών ελέγχων προόδου, την εκτίμηση του πιθανού κόστους των υπηρεσιών του αποθέματος και την ευθύνη της επίλυσης συγκρούσεων μεταξύ των εξωτερικών οντοτήτων που επιδρούν στο απόθεμα. Ακόμα, στις αρμοδιότητες της Διαχείρισης περιλαμβάνεται η θέσπιση διεργασιών για τη διασφάλιση της αξιοποίησης του O AIS στο απόθεμα. Στο λειτουργικό μοντέλο που προτείνεται εδώ η Διαχείριση, εκτός από τα παραπάνω, μπορεί να θεωρηθεί ότι περιλαμβάνει το σύνολο των προαιρετικών σχεδιαστικών επιλογών, οι οποίες, αν υλοποιηθούν, επηρεάζουν εσωτερικά το απόθεμα. Εξαίρεση αποτελούν οι *Εσωτερικές Διαχειριστικές Διεργασίες του ΣΔΒΔ* (φαίνονται με διακεκομμένη γραμμή). Οι επιλογές αυτές αναλυτικότερα είναι:

- *Συμμόρφωση με το O AIS μοντέλο αναφοράς*: Αφορά στη θέσπιση οντοτήτων και διεργασιών, ώστε το ψηφιακό απόθεμα να συμμορφώνεται με τις προδιαγραφές του O AIS συνολικά.
- *Δεικτοδότηση Περιεχομένου Βάσει Οντολογίας*: Τα ψηφιακά αντικείμενα που εισάγονται στο απόθεμα θα δεικτοδοτούνται με βάσει την οντολογία στην οποία ανήκουν. Αυτό σημαίνει ότι για κάθε αντικείμενο του αποθέματος θα παράγεται ένα στιγμιότυπο του οντολογικού σχήματος όπου ανήκει. Τυπικά, με κάθε αντικείμενο θα είναι συσχετισμένο ένα OWL ή DAML+OIL έγγραφο. Τα έγγραφα αυτά στην ιδανική περίπτωση θα δημιουργούνται αυτόματα από τα μεταδεδομένα που παρήχθησαν κατά την Εισαγωγή και από το οντολογικό σχήμα (ή σχήματα) του αποθέματος. Η δεικτοδότηση βάσει οντολογίας θα βοηθήσει αργότερα στην κατανόηση του περιεχομένου από έξυπνους πράκτορες και στην υποβολή έξυπνων ερωτημάτων.
- *Persistency*: Αφορά στην εξασφάλιση της μόνιμης προσπέλασης και της διατήρησης πολλαπλών εκδόσεων των ψηφιακών αντικειμένων. Μπορεί να αποτελέσει εσωτερική ευθύνη του DBMS (τουλάχιστον για την αποθήκευση πολλαπλών εκδόσεων).
- *Εσωτερικές Διαχειριστικές Λειτουργίες του ΣΔΒΔ*: Πρόκειται για τις κλασσικές εσωτερικές λειτουργίες μιας Βάσης Δεδομένων, όπως δεικτοδότηση, οργάνωση του περιεχομένου στο φυσικό μέσο, ER μοντέλο συσχετίσεων κτλ. Αποτελεί σχεδιαστική επιλογή με την έννοια της επιλογής του κατάλληλου συστήματος ΒΔ. Μέρος της καθημερινής διαχείρισης του αποθέματος υλοποιείται από αυτές τις λειτουργίες.

3.4. Κατανάλωση και Πολιτικές Προσπέλασης

Στο επίπεδο της Κατανάλωσης παρέχονται οι επιθυμητές διεπαφές για την κατανάλωση του περιεχομένου του αποθέματος και την διάχυση των υπηρεσιών του προς τις κοινότητες Καταναλωτών. Μεταξύ του επιπέδου του Αποθέματος και της Κατανάλωσης τοποθετείται το ενδιάμεσο υποεπίπεδο των Πολιτικών Προσπέλασης.

Οι Πολιτικές Προσπέλασης στο λειτουργικό μοντέλο αφορούν στην εφαρμογή των προδιαγραφών Ασφάλειας / Πιστοποίησης χρηστών του ψηφιακού αποθέματος. Έτσι, στο υποεπίπεδο αυτό περιλαμβάνεται η υλοποίηση του τμήματος εκείνου της επιχειρησιακής λογικής που θα εξασφαλίζει τον έλεγχο της προσπέλασης του περιεχομένου και των υπηρεσιών του αποθέματος από τους Καταναλωτές. Οι Πολιτικές Προσπέλασης θα πρέπει να επιτρέπουν την απρόσκοπτη προσπέλαση του αποθέματος από την Προσδιορισμένη Κοινότητα σύμφωνα με τους όρους της *Συμφωνίας Παραγγελίας (Order Agreement)*, βλ. OAI & Ορισμό 2). Με άλλα λόγια οι Πολιτικές Προσπέλασης καθιστούν διαθέσιμο όλο ή μέρος του περιεχομένου σε όλους ή συγκεκριμένους χρήστες ή κοινότητες χρηστών, ελέγχουν αν ο χρήστης που αιτείται κατανάλωση πληροί τις προϋποθέσεις για να ικανοποιηθεί το αίτημα του (π.χ. χρέωση) και καθορίζουν αν όλο ή μέρος του περιεχομένου θα είναι διαθέσιμο από όλες ή από συγκεκριμένες διεπαφές Κατανάλωσης.

Οι παρακάτω διεπαφές προτείνονται για το επίπεδο Κατανάλωσης:

- *Ερωτήσεις Χρήστη*: Ο προσδιορισμένος χρήστης του αποθέματος θα μπορεί να θέτει ερωτήματα στις βάσεις δεδομένων του αποθέματος, ως μέσο για την ανάκτηση πληροφορίας και περιεχομένου από αυτό. Στην πράξη οι διεπαφές ερωτήσεων χρήστη υλοποιείται παρέχοντας τη δυνατότητα υποβολής SQL / XQL ερωτημάτων με φιλικό προς το χρήστη τρόπο (πεδία συμπλήρωσης με λέξεις κλειδιά, λογική συσχέτιση μεταξύ των όρων της αναζήτησης –AND, OR...- κτλ). Ως αποτέλεσμα, θα παρουσιάζονται στο χρήστη τόσο τα ψηφιακά αντικείμενα που πληρούν τα κριτήρια της αναζήτησής του, όσο και (όλο ή μέρος από) τα συνοδευτικά μεταδεδομένα τους. Ακόμα ο χρήστης θα είναι σε θέση να αξιοποιεί το σύνολο οντολογιών του αποθέματος και να υποβάλλει *έξυπνα ερωτήματα*, με την έννοια ότι θα μπορεί να ανακτά πληροφορία ακόμα κι αν αυτή δεν είναι ρητά καταχωρημένη στο Απόθεμα. (Για παράδειγμα η ερώτηση: «Βρες μου όλα τα μεταλλικά εκθέματα της Ιστορικής Περιόδου», θα επιστρέφει λ.χ. και τα χάλκινα εκθέματα της ζητούμενης περιόδου, ακόμα κι αν στα μεταδεδομένα τους δεν αναφέρεται ότι το υλικό κατασκευής τους είναι μέταλλο, αφού σε μια οντολογία υλικών η κλάση *Χαλκός* είναι υποκλάση της κλάσης *Μέταλλο*). Εκτός από τη δυνατότητα υποβολής ερωτήσεων, ο χρήστης θα μπορεί να πλοηγείται στο απόθεμα μέσω μιας ιεραρχικής (οντολογικής ή άλλης) οργάνωσης του περιεχομένου.
- *Έξυπνοι Πράκτορες*: Η κατάρτιση οντολογιών για το περιεχόμενο και τις υπηρεσίες του αποθέματος θα καταστήσουν δυνατή την προσπέλαση και κατανάλωσή του από έξυπνους πράκτορες. Οι πράκτορες θα μπορούν να θέτουν έξυπνα ερωτήματα στο απόθεμα και να διαπραγματεύονται την παροχή περιεχομένου και πληροφοριών απ' αυτό, με αυτόματο και διαφανή τρόπο. Ακόμη, οι πράκτορες θα συνεισφέρουν (δεδομένης της διασφάλισης της *πίστης* μεταξύ πρακτόρων-αποθέματος) στην επαύξηση ή τροποποίηση των οντολογιών του.
- *OAI*: Το ψηφιακό απόθεμα θα μπορεί να λειτουργεί ως *Data Provider* (βλ. OAI-PMH, version 2) κάνοντας διαθέσιμα τα μεταδεδομένα του σύμφωνα με το πρωτόκολλο OAI-PMH. Εκτός από το επίπεδο της Κατανάλωσης, το απόθεμα μπορεί να υποστηρίζει το OAI και ως *Service Provider*: θα δέχεται

μεταδεδομένα από άλλους Data Providers (επίπεδο Εισαγωγής) και θα τα καθιστά διαθέσιμα προς κατανάλωση.

- *Z39.50*: Το απόθεμα μπορεί να λειτουργήσει ως server παροχής των πληροφοριών του, υποστηρίζοντας το Z39.50 client / server πρωτόκολλο επικοινωνίας για ανάκτηση πληροφορίας. Η υποστήριξη του Z39.50 θα καταστήσει δυνατή την προσπέλαση του αποθέματος από απομεμακρυσμένα client συστήματα, τα οποία με τη σειρά τους μπορούν να κάνουν διαθέσιμες τις πληροφορίες που ανακτούν στους τελικούς χρήστες τους. Καθώς το Z39.50 είναι ένα διεθνώς τυποποιημένο πρωτόκολλο για ανάκτηση πληροφορίας (ANSI / NISO), η υποστήριξή του θα συνδράμει καταλυτικά στη διαλειτουργικότητα του αποθέματος. Υπάρχουν επίσης και συγκεκριμένα Z39.50 profiles που μπορούν να αξιοποιηθούν από το απόθεμα (π.χ. CIMI Profile για Πληροφορίες Πολιτιστικής Κληρονομιάς). Όπως και στην περίπτωση του OAI, το απόθεμα μπορεί να υποστηρίζει το Z39.50 και ως client, στο επίπεδο Εισαγωγής.
- *UDDI*: Το ψηφιακό απόθεμα μπορεί να ακολουθεί τις προδιαγραφές του UDDI, ώστε να είναι ανακαλύψιμο, προσπελάσιμο και τελικά καταναλώσιμο ως Υπηρεσία Ιστού. Η χρήση της γλώσσας WSDL για την περιγραφή του αποθέματος ως Υπηρεσία Ιστού δεν υπαγορεύεται από τις προδιαγραφές (version 3), ωστόσο ενδείκνυται ως βέλτιστη πρακτική. Ανάλογα με την πολιτική ανάπτυξής του, το απόθεμα μπορεί να λειτουργεί και ως UDDI Node, συγκεντρώνοντας πληροφορίες Υπηρεσιών Ιστού και από άλλα αποθέματα. Ως UDDI Node το απόθεμα μπορεί να αποτελέσει μέρος του ευρύτερου UDDI Business Registry (UBR) ή ακόμα να λειτουργήσει ως Root Node ενός domain-specific registry (π.χ. Registry Αποθεμάτων Πολιτισμικού Περιεχομένου).

3.5. Διαβαθμισιμότητα Λειτουργικού Μοντέλου

Είναι σαφές ότι, για την επίτευξη στοιχειώδους λειτουργικότητας, δεν είναι αναγκαία η υλοποίηση όλων των παραπάνω. Έχει επίσης καταβληθεί προσπάθεια διατήρησης μιας (λεπτής αλήθεια) ισορροπίας μεταξύ του αφηρημένου λειτουργικού σχεδιασμού και της λεπτομερούς τεχνικής υπαγόρευσης της πραγμάτωσής του.

Το λειτουργικό μοντέλο που προτείνεται εδώ έχει σχεδιαστεί έτσι, ώστε να είναι ανοιχτό, ευέλικτο, βαθμωτό και επεκτάσιμο, με νέες προσθήκες. Για παράδειγμα το επίπεδο της Εισαγωγής μπορεί να επαυξηθεί παρέχοντας νέες διεπαφές για εισαγωγή πληροφορίας, όπως υποστήριξη του Z39.50 και του OAI (βλ. και *Παρ. 3.4*). Από την άλλη, στην πορεία υλοποίησης του επιπέδου Κατανάλωσης, δεν χρειάζεται να υλοποιηθούν όλες τις προτεινόμενες διεπαφές, αλλά μπορούν να παραληφθούν ορισμένες ή να υλοποιηθούν σταδιακά ή ακόμα να υποστηριχθούν κάποιες άλλες (όπως η ανάπτυξη και αλληλεπίδραση ενός συστήματος ηλεκτρονικών πωλήσεων με το επίπεδο Κατανάλωσης και το υποεπίπεδο Πολιτικών Προσπέλασης). Ακόμα και στο ιδιαίτερα κρίσιμο επίπεδο Αποθέματος παρέχεται ελαστικότητα και ευελιξία, με μια σειρά σχεδιαστικών επιλογών. Τέλος, οι επιλογές του επιπέδου Διαχείρισης μπορούν να επηρεάσουν σημαντικά την ποιότητα των προσφερόμενων υπηρεσιών του αποθέματος. Σε καμία όμως περίπτωση δεν είναι δεσμευτικές για τη λειτουργικότητά του, παρέχοντας περισσότερο ένα σύνολο προτάσεων για την προσθήκη αξίας (value-add) στο περιεχόμενο και τις υπηρεσίες του αποθέματος.