

Ψηφιακά Δεδομένα

Ψηφιακό Περιεχόμενο

Μεταδεδομένα

Διαλειτουργικότητα

Προδιαγραφές

Ηλεκτρονικές Υποδομές

ΕΠΣΕΤ

Εθνικό Πληροφοριακό
Σύστημα Έρευνας & Τεχνολογίας

SaaS

Πλατφόρμα Παροχής Υπηρεσιών Κατάθεσης,
Διαχείρισης & Διάθεσης Ανοικτών Δημόσιων
Δεδομένων Τεκμηρίωσης & Ψηφιακού Περιεχομένου

Προδιαγραφές και χαρακτηριστικά διαλειτουργικότητας
για ανοικτό ψηφιακό περιεχόμενο

EKT

ΕΘΝΙΚΟ ΚΕΝΤΡΟ
ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL
DOCUMENTATION
CENTRE



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης



Ψηφιακή Ελλάδα
"Όλα είναι δυνατόν"
"Ποιότητα Ζωής για όλους"



ΕΣΠΑ
2007-2013
Πρόγραμμα για την ανάπτυξη
Ποιότητα ζωής για όλους



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Παιδείας και Θρησκευμάτων,
Πολιτισμού και Αθλητισμού

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής
Ένωσης - Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης

Προδιαγραφές και χαρακτηριστικά διαλειτουργικότητας για ανοικτό ψηφιακό περιεχόμενο

Άξονας Προτεραιότητας:	ΤΠΕ και Βελτίωση της Ποιότητας Ζωής
Στόχος:	Σύγκλιση
Τίτλος Πράξης:	«Εθνικό Πληροφοριακό Σύστημα Έρευνας και Τεχνολογίας/Κοινωνικά Δίκτυα – Περιεχόμενο Παραγόμενο από Χρήστες»
Κωδικός Πράξης (ΟΠΣ):	296115
Τίτλος Πράξης:	«Πλατφόρμα Παροχής Υπηρεσιών Κατάθεσης, Διαχείρισης και Διάθεσης Ανοιχτών Δεδομένων και Ψηφιακού Περιεχομένου»
Κωδικός Πράξης (ΟΠΣ):	327378

Αρχική έκδοση: Μάιος 2009

1^η Επικαιροποίηση: Νοέμβριος 2011

2^η Επικαιροποίηση: Μάρτιος 2012

3^η Επικαιροποίηση: Μάρτιος 2013

Συντάκτες: Δρ Παναγιώτης Σταθόπουλος
Δρ Νίκος Χούσος
Γιώργος Σταύρου

Copyright © 2012-2013 Εθνικό Κέντρο Τεκμηρίωσης / Εθνικό Ίδρυμα Ερευνών
Βασιλέως Κωνσταντίνου 48, 11635 Αθήνα

Τηλ.: 210 7273900-02 • Fax: 210 7246824

e-mail: ekt@ekt.gr • <http://www.ekt.gr>

Το έργο αυτό διατίθεται με άδεια Creative Commons

Αναφορά - Μη-Εμπορική Χρήση - Όχι Παράγωγα Έργα 3.0 Ελλάδα

Προκειμένου να δείτε αντίγραφο της άδειας επισκεφθείτε:

<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.el>

Μόνιμη ηλεκτρονική διεύθυνση για αναφορά: <http://hdl.handle.net/10442/8887>

Περιεχόμενα

Εισαγωγή.....	3
1. Σκοπός	5
2. Αποθετήρια ανοικτού ψηφιακού περιεχομένου	7
3. Προδιαγραφές για την ψηφιοποίηση, οργάνωση και διάθεση ανοικτού ψηφιακού περιεχομένου.....	9
3.1. Ψηφιοποίηση	9
3.2. Μεταδεδομένα και διαλειτουργικότητα	11
3.3. Μόνιμοι προσδιοριστές (διευθύνσεις) πόρων ανοικτού ψηφιακού περιεχομένου	19
3.4. Προδιαγραφές για διάθεση ανοικτού ψηφιακού περιεχομένου	20
3.5. Διαλειτουργικότητα με καταλόγους και e-υποδομή ανοικτού ψηφιακού περιεχομένου ΕΚΤ	22
4. Ενδεικτικό λογισμικό για την ανάπτυξη υποδομών διάθεσης ανοικτού ψηφιακού περιεχομένου.....	24
5. Σύνοψη λειτουργικότητας χαρακτηριστικών και προδιαγραφών	26
Βιβλιογραφία	28
Κατάλογος Πινάκων και Εικόνων.....	30
Πίνακες.....	30
Εικόνες	30

Εισαγωγή

Στη μελέτη αυτή αναπτύσσονται οι προδιαγραφές καθώς και τα χαρακτηριστικά διαλειτουργικότητας για την ανάπτυξη ανοικτού ψηφιακού περιεχομένου, όπως προκύπτουν από την εμπειρία του ΕΚΤ και από το διεθνές περιβάλλον και τα πρότυπα στις αντίστοιχες περιοχές. Οι προδιαγραφές αυτές διέπουν συνολικά την υλοποίηση σημαντικών δράσεων που αναπτύσσει το ΕΚΤ, όπως η πράξη «Εθνικό Πληροφοριακό Σύστημα Έρευνας και Τεχνολογίας/Κοινωνικά Δίκτυα – Περιεχόμενο Παραγόμενο από Χρήστες» και η πράξη «Πλατφόρμα Παροχής Υπηρεσιών Κατάθεσης, Διαχείρισης και Διάθεσης Ανοιχτών Δεδομένων και Ψηφιακού Περιεχομένου», οι οποίες αξιοποιούν πόρους από το Επιχειρησιακό Πρόγραμμα «Ψηφιακή Σύγκλιση» και τα Περιφερειακά Επιχειρησιακά Προγράμματα (ΕΣΠΑ 2007-2013).

Η αρχική έκδοση της μελέτης και η 1^η και 2^η επικαιροποίησή της έλαβαν χώρα στο πλαίσιο του Έργου «Εθνικό Πληροφοριακό Σύστημα Έρευνας και Τεχνολογίας/Κοινωνικά Δίκτυα – Περιεχόμενο Παραγόμενο από Χρήστες», ενώ η 3^η επικαιροποίηση υλοποιήθηκε στο πλαίσιο του έργου «Πλατφόρμα Παροχής Υπηρεσιών Κατάθεσης, Διαχείρισης και Διάθεσης Ανοιχτών Δεδομένων και Ψηφιακού Περιεχομένου».

Το Εθνικό Κέντρο Τεκμηρίωσης αποτελεί επιστημονική εγκατάσταση εθνικής χρήσης στο Εθνικό Ίδρυμα Ερευνών. Με τον όρο «επιστημονική εγκατάσταση εθνικής χρήσης» αποτυπώνεται ο ιδιαίτερος χαρακτήρας του ΕΚΤ ως διακριτής εθνικής υποδομής, με θεσμικό ρόλο τη συλλογή, οργάνωση και διάχυση της επιστημονικής και τεχνολογικής πληροφορίας, εντός και εκτός της χώρας. Με διαρκή παρουσία στην ελληνική επιστημονική κοινότητα από το 1980, το ΕΚΤ αξιοποιεί τις Τεχνολογίες Πληροφορίας και Επικοινωνιών, σε συνδυασμό με σύγχρονες επιχειρησιακές μεθόδους, ώστε να αναπτύσσει καινοτόμες δράσεις, προσανατολισμένες στην ενίσχυση της πρόσβασης στο ψηφιακό περιεχόμενο. Για τον σκοπό αυτό, συνεργάζεται με σημαντικούς φορείς παραγωγής έγκριτου περιεχομένου, όπως βιβλιοθήκες, αρχεία, μουσεία, ερευνητικά κέντρα.

Μέσα από τις δράσεις που αναπτύσσει το ΕΚΤ, έχει διαπιστωθεί ότι αποτελεί στρατηγικής σημασίας προτεραιότητα να σχεδιαστούν κεντρικά και να παρασχεθούν ενιαία, νέα εργαλεία και υπηρεσίες, βασισμένα στις τεχνολογίες του παγκόσμιου ιστού για την **οργάνωση και διάθεση του περιεχομένου και των δεδομένων**.

Στη κατεύθυνση αυτή, το ΕΚΤ υλοποιεί την πράξη «Εθνικό Πληροφοριακό Σύστημα Έρευνας και Τεχνολογίας (ΕΠΣΕΤ)» αναπτύσσοντας μια **εθνική ερευνητική e-υποδομή για την οργάνωση και διάθεση ψηφιακής πληροφορίας και περιεχομένου επιστήμης, τεχνολογίας και πολιτισμού**. Η υλοποίηση της πρώτης φάσης του ΕΠΣΕΤ ξεκίνησε το 1996, με βάση τη μοναδική τεχνογνωσία και εμπειρία του Εθνικού Κέντρου Τεκμηρίωσης, τα υπάρχοντα και αναδυόμενα διεθνή πρότυπα και τις πρωτοποριακές Τεχνολογίες Πληροφορίας και Επικοινωνιών.

Η e-υποδομή του ΕΚΤ ενσωματώνει τεχνολογίες αιχμής, αξιοποιεί μακροχρόνιες συνεργασίες με φορείς έρευνας, παιδείας και πολιτισμού, εφαρμόζει διεθνή πρότυπα σε όλα τα επίπεδα (οργάνωση δεδομένων, διαφύλαξη περιεχομένου, παροχή υπηρεσιών, διαλειτουργικότητα συστημάτων) και υλοποιεί την πολιτική της Ανοικτής Πρόσβασης στα ερευνητικά αποτελέσματα. Εξυπηρετεί καταγεγραμμένες ανάγκες χρηστών, υποστηρίζει τη μεταφορά γνώσης και εντάσσεται στο υπό διαμόρφωση διεθνές δίκτυο ανάλογων υποδομών.

Ένα ακόμη αναπτυξιακό εργαλείο του ΕΚΤ αποτελεί το έργο «Πλατφόρμα Παροχής Υπηρεσιών Κατάθεσης, Διαχείρισης και Διάθεσης Ανοιχτών Δεδομένων και Ψηφιακού Περιεχομένου» που υλοποιεί στο πλαίσιο του Επιχειρησιακού Προγράμματος «Ψηφιακή Σύγκλιση» (ΕΣΠΑ 2007-2013), με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης.

Αντικείμενο του έργου είναι **η ανάπτυξη και παροχή σύγχρονων ψηφιακών υπηρεσιών σε φορείς παραγωγής ψηφιακού περιεχομένου** (βιβλιοθήκες, μουσεία, αρχεία και πολιτιστικούς φορείς ευρύτερα). Οι υπηρεσίες αυτές αποσκοπούν στην ενίσχυση της ψηφιακής παρουσίας των φορέων, μέσα από την αξιοποίηση ολοκληρωμένων λύσεων που επιτρέπουν την **τεκμηριωμένη απόθεση, ασφαλή διαφύλαξη, οργάνωση και διάθεση του ψηφιακού τους περιεχομένου**.

Οι υπηρεσίες του έργου βασίζονται σε τεχνολογίες αιχμής, όπως το μοντέλο SaaS (Software as a Service=Λογισμικό ως Υπηρεσία) και το Cloud Computing (Υπολογιστικό Νέφος) και παρέχονται από το ΕΚΤ **χωρίς κόστος για τους φορείς**, διαδικτυακά, χωρίς τοπική εγκατάσταση, αλλά **προσαρμοσμένες στις απαιτήσεις του κάθε φορέα**. Οι υπηρεσίες αυτές συνοδεύονται από υπηρεσίες εκπαίδευσης και υποστήριξης των φορέων στους οποίους απευθύνονται, ώστε να διασφαλισθεί η αποτελεσματική αξιοποίησή τους.

Στόχος της παρέμβασης αυτής είναι η ανάπτυξη ενός ισχυρού δικτύου φορέων, το οποίο θα συνδιαμορφώσει τη στρατηγική και τις προδιαγραφές για την αύξηση του έγκριτου ψηφιακού περιεχομένου της χώρας και την προώθηση της επανάχρησής του.

1. Σκοπός

Τα έργα ψηφιοποίησης και διάθεσης περιεχομένου των προηγούμενων προγραμματικών περιόδων είχαν σημαντική συνεισφορά στην ψηφιοποίηση του Ελληνικού επιστημονικού και πολιτιστικού αποθέματος. Ωστόσο, με βάση την προηγούμενη εμπειρία ανέκυψαν μια σειρά από σημαντικά ζητήματα κατά την υλοποίηση τους που:

- περιορίζουν τη δυναμική του παραχθέντος περιεχομένου, μειώνουν την αξία της επένδυσης και των δυνατοτήτων ευρείας επαναχρησιμοποίησης και οργάνωσης του περιεχομένου που έχει πραγματοποιηθεί.
- δεν επιτρέπουν την σε ευρεία κλίμακα εφαρμογή των ραγδαίων τεχνολογικών εξελίξεων που χαρακτηρίζουν το περιβάλλον διάθεσης ανοικτών τεκμηριωμένων δημοσίων δεδομένων και ανοικτού τεκμηριωμένου ψηφιακού περιεχομένου, σύμφωνα και με τις απαιτήσεις για την Ηλεκτρονική Διακυβέρνηση.
- δεν λαμβάνουν υπόψη νέες καινοτομικές προσεγγίσεις διάθεσης εφαρμογών και ολοκληρωμένων υπηρεσιών, όπως είναι υποδομές Cloud γενικά και υποδομές παροχής Software as a Service, ειδικότερα, για τη διάθεση εφαρμογών με την μορφή υπηρεσίας όπως προσδιορίζεται και στο Ψηφιακό Θεματολόγιο 2020 της Ευρωπαϊκής Επιτροπής.

Στο λειτουργικό επίπεδο προβλήματα που έχουν αναγνωριστεί αφορούν στην:

- πολυδιάσπαση των συστημάτων που χρησιμοποιούνται για την διάθεση του περιεχομένου,
- το αυξημένο κόστος συντήρησης και λειτουργίας τους σε παραγωγική φάση ανά μεμονωμένη εγκατάσταση,
- τις μειωμένες δυνατότητες επανάχρησης του περιεχομένου, τεχνογνωσίας και συστημάτων,
- την αδυναμία αποτελεσματικής του ένταξης σε διεθνείς καταλόγους ή και σε μηχανές αναζήτησης,
- την μειωμένη διαθεσιμότητα και επίπεδο ασφάλειας ανά μεμονωμένη εγκατάσταση,
- την έλλειψη μηχανισμών για την εξασφάλιση της μακροχρόνιας διατήρησης, αξιοποίησης και ασφάλειας του περιεχομένου.

Μέρος των παραπάνω προβλημάτων οφείλονται και στις ιδιαιτερότητες του ελληνικού περιβάλλοντος, το οποίο αν και χαρακτηρίζεται από σημαντικό αριθμό φορέων οι οποίοι διαθέτουν πολύτιμο περιεχόμενο που μπορεί να ικανοποιήσει διευρυμένες απαιτήσεις χρήσης, στις περισσότερες περιπτώσεις δεν διαθέτουν το απαιτούμενο, τεχνολογικά κρίσιμο, μέγεθος για την ποιοτική και διατηρήσιμη διάθεση σύνθετου ψηφιακού περιεχομένου.

Κατά την προηγούμενη προγραμματική περίοδο είχαν δοθεί στους αναδόχους έργων με τη μορφή μελετών ^{[1][2][3]} καλές πρακτικές και κατευθυντήριες γραμμές για την αντιμετώπιση τεχνολογικών θεμάτων, ωστόσο, με βάση την εμπειρία της υλοποίησης των έργων φάνηκε ότι οι τεχνολογικές κατευθύνσεις δεν αρκούν από μόνες τους για την αντιμετώπιση των προαναφερθέντων ζητημάτων, ιδιαίτερα σε μια σειρά από κρίσιμα στοιχεία διαλειτουργικότητας, δυνατοτήτων επαναξιοποίησης περιεχομένου και εμπειρίας του χρήστη.

Πέρα λοιπόν από την επικαιροποίηση επιμέρους στοιχείων των υπάρχουσών μελετών, καθώς και την προσθήκη των νέων, σε τεχνολογικό και επιχειρησιακό επίπεδο, θεμάτων που έχουν προκύψει, είναι απαραίτητο να διαμορφωθεί ένα πλαίσιο για την τήρηση των προδιαγραφών που πρέπει να πληρούν τα συστήματα διάθεσης ανοικτού ψηφιακού περιεχομένου, συνδυαζόμενο με τα απαραίτητα

σημεία ελέγχου διαλειτουργικότητας, ώστε να αποφευχθούν αστοχίες κατά τη διάρκεια υλοποίησης των έργων, να μειωθεί το κόστος τους και να ληφθεί πρόνοια για τη βιωσιμότητα και μακροχρόνια διατηρησιμότητα του περιεχομένου που παράγεται.

Για τον σκοπό αυτό, στην παρούσα μελέτη, έχουν προδιαγραφεί οι κύριες λειτουργικές προδιαγραφές και οι καλές πρακτικές που θα πρέπει να υποστηρίζονται, κατά ελάχιστο από το σύστημα ψηφιακού αποθετηρίου που χρησιμοποιείται για την οργάνωση, απόθεση και διάθεση του ανοικτού ψηφιακού περιεχομένου. Για τις κρίσιμες προδιαγραφές διαλειτουργικότητας επιπλέον αξία έχει και ο έλεγχος στις απαιτήσεις και με αυτοματοποιημένα εργαλεία ελέγχου συμμόρφωσης.

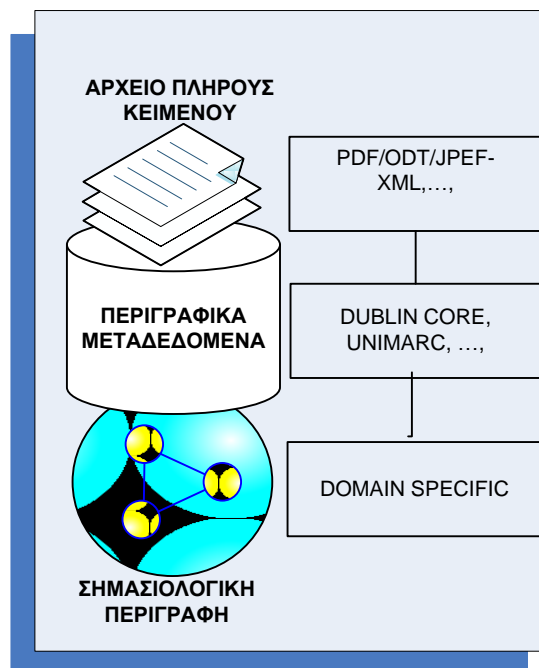
Σκοπός του παρόντος κειμένου είναι να ορίσει ένα σαφές και ρεαλιστικό πλαίσιο για την επίτευξη διαλειτουργικότητας του συνόλου των συστημάτων και του περιεχομένου που θα προκύψουν από έργα που έχουν ανάμεσα στους κύριους στόχους τους την διάθεση ανοικτού ψηφιακού περιεχομένου. Το πλαίσιο αυτό δίνει τη δυνατότητα για περαιτέρω επέκταση ώστε να καλύπτονται οι ανάγκες επιπλέον εξειδικευμένων κατηγοριών περιεχομένου καθώς και να ενσωματώνονται οι διαρκείς διεθνείς εξελίξεις σε θέματα τεχνολογίας, προτυποποίησης και υποδομών.

Το πλαίσιο και οι αντίστοιχες προδιαγραφές έχουν προκύψει με βάση την πολύχρονη εθνική και διεθνή εμπειρία του Εθνικού Κέντρου Τεκμηρίωσης, την υλοποίηση της εθνικής e-υποδομής ανοικτού ψηφιακού περιεχομένου «Εθνικό Πληροφοριακό Σύστημα Έρευνας και Τεχνολογίας» (<http://www.epset.gr>) αναγνωρίζοντας ως βάση τις διεθνείς πρακτικές, σε επίπεδο ευρωπαϊκών κρατών (Ηνωμένο Βασίλειο, Ολλανδία, Γαλλία, Σουηδία κ.α.) και συνολικά της ΕΕ (προγράμματα e-content, ICT PSP, e-infrastructure κ.α.), καλά παραδείγματα επιμέρους διεθνών αποθετηρίων και συστημάτων ψηφιακού περιεχομένου μεγάλης κλίμακας (Open Library του Internet Archive, Google Books και Art Project, κ.α.) καθώς και τις εξελίξεις, ευκαιρίες και δυνατότητες που προσφέρουν SaaS (Software as a Service Cloud) υποδομές για ψηφιακό περιεχόμενο. Το Εθνικό Κέντρο Τεκμηρίωσης, έχει θεσμικό ρόλο στην επιστημονική πληροφόρηση και μακροχρόνια παρουσία στη διάθεση ανοικτού ψηφιακού περιεχομένου και υπηρεσιών στην ακαδημαϊκή, ερευνητική και επιστημονική κοινότητα.

2. Αποθετήρια ανοικτού ψηφιακού περιεχομένου

Ως αποθετήρια (repositories) ορίζονται συστήματα απόθεσης, οργάνωσης και διάθεσης του οργανωμένου περιεχομένου που σκοπό έχουν τη συλλογή, διάθεση και μακροχρόνια διατήρηση ψηφιακού περιεχομένου, προερχόμενου είτε από ψηφιοποίηση είτε από υλικό που έχει δημιουργηθεί εξ αρχής σε ηλεκτρονική μορφή (born-digital). Διαφοροποιούνται από απλά συστήματα διαχείρισης περιεχομένου, λόγω της κρίσιμης σημασίας που έχουν για την λειτουργία τους μιας σειράς παραγόντων διαλειτουργικότητας, επαναχρησιμοποίησης και διατήρησης των δεδομένων και των πλούσιων γενικά μεταδεδομένων του ψηφιακού περιεχομένου.

Τα αποθετήρια περιεχομένου στο σύνολό τους δομούνται γύρω από το ψηφιακό τεκμήριο, το οποίο αποτελείται από το σώμα του κυρίως τεκμηρίου (πλήρες κείμενο, εικόνα, κ.α.), τα περιγραφικά μεταδεδομένα και τη σημασιολογική τους ερμηνεία που τα συνοδεύει. Η δομημένη αυτή πληροφορία υποστηρίζεται από μια σειρά νέων αλλά και ώριμων ανοικτών διεπαφών, σχημάτων μεταδεδομένων και σημασιολογικών περιγραφών για την ανταλλαγή, αναζήτηση, συσσώρευση και διασύνδεση του περιεχομένου με οργανωμένο τρόπο, οι οποίες έχουν προκύψει κατά τα τελευταία χρόνια σε διεθνές επίπεδο.



Εικόνα 1 Απλουστευμένη μορφή ψηφιακού τεκμηρίου και σχετικών προτύπων

Βάσει του περιεχομένου που διαχειρίζονται, τα αποθετήρια μπορούν να διακριθούν συνήθως σε τρεις κατηγορίες:

- Αποθετήρια / συλλογές ψηφιακού πολιτιστικού ή/ και ιστορικού αποθέματος (π.χ. εικόνες έργων τέχνης, αρχαιολογικών μνημείων, χειρογράφων).
- Αποθετήρια επιστημονικού περιεχομένου (π.χ. επιστημονικές δημοσιεύσεις όπως άρθρα, βιβλία, πρακτικά συνεδρίων).
- Αρχεία διάφορων τύπων και μορφών (π.χ. προσωπικά αρχεία προσωπικοτήτων, οπτικοακουστικά αρχεία ραδιοτηλεοπτικών οργανισμών, αρχεία περιοδικών εκδόσεων).

Οι επιμέρους μορφές περιεχομένου, τρόποι οργάνωσης και συγκεκριμένα πρότυπα ποικίλουν ανάλογα με το είδος, μοιράζονται ωστόσο τις βασικές αρχές διαλειτουργικότητας και ανάγκες παρουσίασης και διάθεσης του περιεχομένου.

Σημαντικό ρόλο στην εξάπλωση των δυνατοτήτων χρήσης, αναζήτησης και οργάνωσης του περιεχομένου διαδραματίζουν επίσης οι συσσωρευτές μεταδεδομένων, οι οποίοι συλλέγουν μεταδεδομένα από τα επιμέρους αποθετήρια με ανοικτές διεπαφές. Οι συσσωρευτές αυτοί μπορούν να λειτουργήσουν ταυτόχρονα και ως «μητρώο» του ανοικτού ψηφιακού περιεχομένου, όπου καταγράφεται η προέλευση του, η τοποθεσία μόνιμης αναφοράς και διάθεσης του, συγκεντρωτικά στατιστικά στοιχεία κ.α. Στην συνέχεια και για λόγους ευκολίας οι έννοιες «συσσωρευτής μεταδεδομένων» και «Μητρώο» ανοικτού ψηφιακού περιεχομένου θα χρησιμοποιούνται εναλλάξιμα.

Στην συνέχεια αναφέρονται οι βασικές απαιτήσεις προδιαγραφών και διαλειτουργικότητας, σε κάθε επίπεδο πρόσκτησης, επεξεργασίας και διάθεσης του ψηφιοποιημένου ή ψηφιακού περιεχομένου.

3. Προδιαγραφές για την ψηφιοποίηση, οργάνωση και διάθεση ανοικτού ψηφιακού περιεχομένου

3.1. Ψηφιοποίηση

Οι διαδικασίες, τα πρότυπα και οι μέθοδοι ψηφιοποίησης του υλικού ενδιαφέροντος έχουν αντιμετωπιστεί, με σημαντικό βαθμό λεπτομέρειας, στο αντίστοιχο τμήμα της μελέτης με τίτλο «Οδηγός καλών πρακτικών για την ψηφιοποίηση και τη μακροπρόθεσμη διατήρηση του πολιτιστικού περιεχομένου» [1], και ενσωματώνουν υλικό από αντίστοιχες διαδεδομένες διεθνείς μελέτες και πηγές [4][5][6][7]. Ιδιαίτερη σημασία έχουν οι ελάχιστες απαιτήσεις ποιότητας που αναφέρονται στην μελέτη [1], οι οποίες αναπαράγονται στη συνέχεια:

Πρωτότυπο αντικείμενο	Ελάχιστη ανάλυση	Χρωματικό βάθος
Φωτοτυπημένο υλικό (ασπρόμαυρο)	200-300 dpi	8 bit γκρι
Έντυπο υλικό (ασπρόμαυρο)	400 dpi ή 4000 pixels στη μεγαλύτερη διάσταση	8 bit γκρι
Έντυπο υλικό (έγχρωμο)	400 dpi ή 4000 pixels στη μεγαλύτερη διάσταση	24 bit
Χάρτες και γραφικά (ασπρόμαυρα)	300 dpi ή 4000 pixels στη μεγαλύτερη διάσταση	8 bit γκρι
Χάρτες και γραφικά (έγχρωμα)	300 dpi ή 4000 pixels στη μεγαλύτερη διάσταση	24 bit
Φωτογραφίες (ασπρόμαυρες)	600 dpi ή 5000 pixels στη μεγαλύτερη διάσταση	8 bit γκρι
Φωτογραφίες (έγχρωμες)	600 dpi ή 5000 pixels στη μεγαλύτερη διάσταση	24 bit
Έργα τέχνης (ασπρόμαυρα)	600 dpi ή 5000 pixels στη μεγαλύτερη διάσταση	8 bit γκρι
Έργα τέχνης, υφάσματα (έγχρωμα)	600 dpi ή 5000 pixels στη μεγαλύτερη διάσταση	24 bit
35mm slides, αρνητικά κλπ (ασπρόμαυρα)	2400 dpi	8 bit γκρι
35mm slides, αρνητικά κλπ (έγχρωμα)	2400 dpi	24 bit
6cm X 6cm slides (ασπρόμαυρα)	2000 dpi	8 bit γκρι
6cm X 6cm slides (έγχρωμα)	2000 dpi	24 bit
Slides ή πλάκες από γυαλί (ασπρόμαυρα)	600 dpi	8 bit γκρι

Πίνακας 1. Ελάχιστες απαιτήσεις για ψηφιοποίηση διαφορετικών αντικειμένων σύμφωνα με [1]

Ταυτόχρονα, στη μελέτη επισημαίνονται μια σειρά από, αυτονόητες μεν αλλά όχι πάντα τηρούμενες, τεχνικές προδιαγραφές που επηρεάζουν την ποιότητα του τελικώς παραγόμενου υλικού και την τελική εμπειρία του χρήστη (π.χ. διόρθωση ψηφιακών εικόνων, ευθυγράμμιση, crop, κ.α.). Συνολικά, οι προδιαγραφές που σημειώνονται στη μελέτη ^[1] εφόσον λαμβάνονται υπόψη **είναι επαρκείς**, αν και πρέπει να εξεταστεί η επικαιροποίηση τους όπως π.χ με την προσθήκη της δυνατότητας για ευρεία αποθήκευση των εικόνων και σε μορφή lossless JPEG 2000 κ.α..

Ωστόσο, επιπλέον της τήρησης των παραπάνω προδιαγραφών απαιτείται η εφαρμογή **επεξεργασίας, στην φάση της ψηφιοποίησης, με σύστημα οπτικής αναγνώρισης χαρακτήρων (OCR) για έντυπο υλικό, το οποίο αποτελεί το κύριο τεκμήριο ή το συνοδευτικό υλικό τεκμηρίωσης ενός αντικειμένου**, ώστε να γίνεται δυνατή η αναζήτηση από μηχανές αναζήτησης στο διαδίκτυο, το Αποθετήριο του φορέα, και συσσωρευτές ανοικτού ψηφιακού περιεχομένου, η αναζήτηση και το indexing στο πλήρες κείμενο (full text search).

Προτείνεται να απαιτείται τουλάχιστον «αδιόρθωτο», όπως ονομάζεται, OCR για κείμενα Νέων Ελληνικών¹ και άλλων διαδεδομένων γλωσσών εφόσον πρόκειται για ξενόγλωσσο υλικό, ως η προσέγγιση που εξασφαλίζει τον καλύτερο λόγο αποτελεσμάτων προς κόστος. Η απαίτηση αυτή είναι κρίσιμη ώστε μετά τη διαδικασία της ψηφιοποίησης το ψηφιακό υλικό να είναι **αναζητήσιμο στο πλήρες κείμενο** στο αποθετήριο, στο διαδίκτυο και στο μητρώο ανοικτού ψηφιακού περιεχομένου, εφόσον είναι διαθέσιμα εκεί εκτός των μεταδεδομένων και τα δεδομένα οπτικής αναγνώρισης χαρακτήρων που έχουν προκύψει από την διαδικασία OCR.

Για την επεξεργασία των ψηφιοποιημένων αρχείων για το σύνολο των έργων ανοικτού ψηφιακού περιεχομένου με λογισμικό OCR υπάρχουν οι παρακάτω δυνατότητες:

- A. **Η απαίτηση επεξεργασίας OCR από τους ανάδοχους των έργων ψηφιοποίησης.** Αποτελεί την πλέον ενδεδειγμένη λύση, μιας και η επιπλέον οικονομική επιβάρυνση είναι αναλογικά μικρή σε σχέση με το συνολικό κόστος (σύμφωνα με εμπειρικές εκτιμήσεις από 0,01 έως και 0,02€/σελίδα). Τα παραδοτέα από την επεξεργασία OCR ανά τεκμήριο θα πρέπει να είναι τα εξής:
- Το κείμενο σε ξεχωριστό αρχείο τύπου .txt, κωδικοποίησης UTF-8. Στο αρχείο κειμένου πρέπει να περιλαμβάνονται οι αλλαγές γραμμής και σελίδας.
 - Το κείμενο σε μορφή «Image PDF with hidden text» που επιτρέπει την αναζήτηση και επιλογή κειμένου σε ξεχωριστό αρχείο PDF, το οποίο συντίθεται από τις σαρωμένες εικόνες και το κείμενο που προκύπτει από το OCR.
 - Αναλυτικό παραγόμενο OCR κείμενο σε μορφή ABBYY XML ή hOCR, που περιλαμβάνει στοιχεία θέσης για κάθε χαρακτήρα / λέξη του κειμένου και επιτρέπει την παρουσίαση του κειμένου με εφαρμογή φυλλομέτρησης / ανάγνωση ως e-book (εναλλακτική του ανοίγματος ως PDF αρχείο) με αναζήτηση και hit highlighting.
- B. Σε περίπτωση που η ψηφιοποίηση διενεργείται εσωτερικά από τον τελικό δικαιούχο χωρίς την προσφυγή σε εξωτερικούς αναδόχους, και για ήδη υπάρχων υλικό προτείνεται η **αξιοποίηση κεντρικής υποδομής OCR για μαζική οπτική αναγνώριση κειμένων** του Εθνικού Κέντρου Τεκμηρίωσης.
- Γ. Η εναλλακτική λύση της τοπικής επεξεργασίας OCR μεμονωμένα για κάθε τελικό δικαιούχο, κρίνεται βιώσιμη μόνο εφόσον πρόκειται για φορείς που η ψηφιοποίηση αποτελεί μια συνεχόμενη και διαρκή διαδικασία.

¹ Αφορά κείμενα από τυπογραφικά δοκίμια και όχι χειρόγραφα κείμενα.

Οι παραπάνω απαιτήσεις δίνονται στον κάτωθι πίνακα (Πίνακας 2) σε μορφή κατάλληλη για συμπερίληψη σε πίνακες συμμόρφωσης.

A/A	Απαιτηση	Περιγραφή
1.	Οπτική Αναγνώριση Χαρακτήρων (OCR) για υποστήριξη αναζήτησης και ευρετηρίασης του πλήρους κειμένου	Παραγωγή αδιόρθωτου OCR για το σύνολο του προς ψηφιοποίηση έντυπου περιεχομένου
2.	Μορφή αποτελεσμάτων OCR	Παράδοση αρχείων text και XML σε μορφή UTF8 αδιόρθωτου OCR. Το text αρχείο πρέπει να βρίσκεται ξεχωριστό αρχείο τύπου κωδικοποίησης UTF-8 και να συμπεριλαμβάνει αλλαγές γραμμής και σελίδας.
3.	Μορφή αποτελεσμάτων OCR	Σε αναλυτικό αρχείο στις πρότυπες μορφές μορφή ABBYY XML ή την ανοικτή μορφή hOCR, που περιλαμβάνουν στοιχεία θέσης για κάθε χαρακτήρα / λέξη του κειμένου και επιτρέπουν την παρουσίαση του κειμένου με εφαρμογή φυλλομέτρησης και με δυνατότητες αναζήτησης με hit highlighting.
4.	Παράδοση τελικού αρχείου	Παράδοση αρχείων σε μορφή PDF, ενσωμάτωση στο PDF αρχείο «Image PDF with hidden text» που επιτρέπει την αναζήτηση και επιλογή κειμένου

Πίνακας 2. Επιπρόσθετες απαιτούμενες προδιαγραφές ψηφιοποίησης

3.2. Μεταδεδομένα και διαλειτουργικότητα

Ιδιαίτερη συνάφεια με τη διαλειτουργικότητα σε επίπεδο μεταδεδομένων έχουν οι έννοιες της διαλειτουργικότητας και των ανοικτών δεδομένων, οι οποίες ορίζονται στην πρόσφατη [Δημόσια διαβούλευση για τις γενικές αρχές και προτεραιότητες Σχεδιασμού του Ε.Π. «Ψηφιακή Σύγκλιση»](#) [8]:

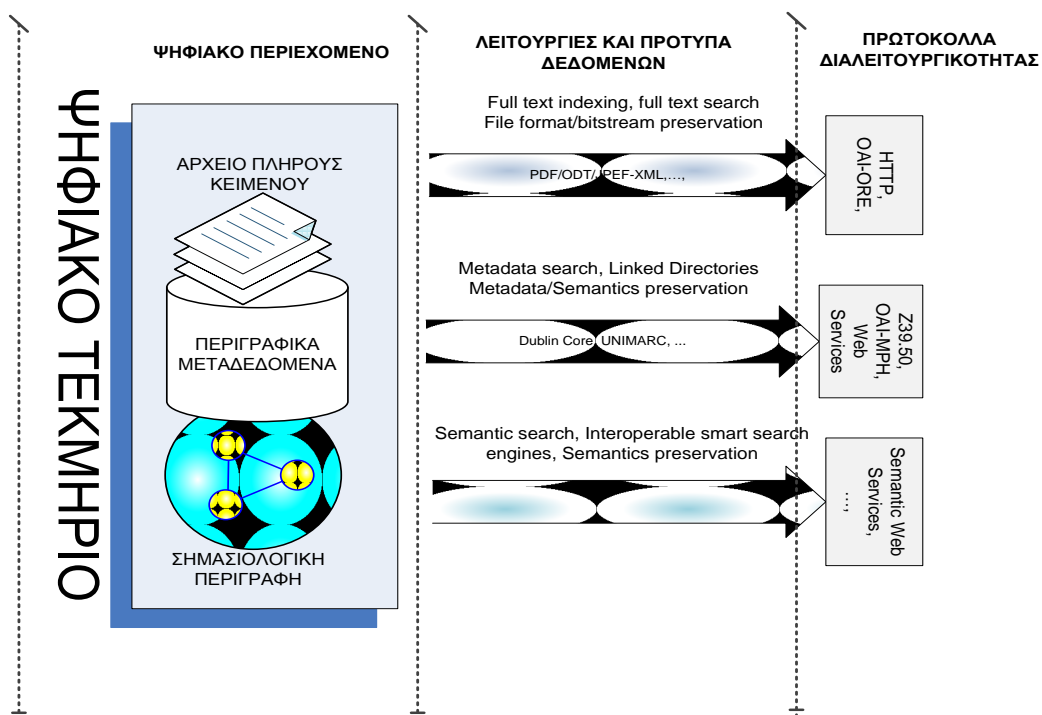
- **Διαλειτουργικότητα** που εξασφαλίζεται από διεπαφές που επιτρέπουν τη συνεργασία μεταξύ διαφορετικών οργανισμών και εφαρμογών (ανοικτά APIs και Web Services, πλήρης υιοθέτηση προσέγγισης Web2.0). Σκοπός είναι η δημιουργία εφαρμογών που θα διαθέτουν τα δεδομένα τους σε μορφή αξιοποιήσιμη από κάθε ενδιαφερόμενο –πολίτη ή φορέα-, προσφέροντας έτσι τη δυνατότητα εκμετάλλευσής τους από άλλες εφαρμογές, χωρίς την ανάγκη σύνθετων λύσεων ολοκλήρωσης ετερογενών συστημάτων.
- **Ανοικτά δεδομένα** με ελεύθερη πρόσβαση από το σύνολο των υπηρεσιών, των επιχειρήσεων και των πολιτών. Ένας από τους σκοπούς του Προγράμματος «Ψηφιακή Σύγκλιση» είναι η διάχυση της πληροφορίας, και κατά κύριο λόγο του πολιτιστικού/γνωσιακού πλούτου της χώρας που υπάρχει ή μετατρέπεται σε ψηφιακή μορφή. Τα δεδομένα των εντασσόμενων έργων θα πρέπει να είναι **ολοκληρωμένα, σωστά κατατεμημένα, προσβάσιμα και δεκτικά επεξεργασίας από Η/Υ**.

Ειδικότερα στα συστήματα αποθετηρίων/ ψηφιακών βιβλιοθηκών, η διαλειτουργικότητα στο πεδίο των μεταδεδομένων μπορεί να θεωρηθεί πως έχει τις παρακάτω τρεις διαστάσεις [9]:

- **Διαλειτουργικότητα σε επίπεδο συστημάτων αποθετηρίων και μητρώου ανοικτού περιεχομένου.** Επιτρέπει την online διάθεση των μεταδεδομένων ενός αποθετηρίου σε τρίτες εφαρμογές και συστήματα. Το αποθετήριο παρέχει ανά πάσα στιγμή τα μεταδεδομένα του μέσω μιας προγραμματιστικής διεπαφής που υλοποιείται βάσει ενός συγκεκριμένου πρωτοκόλλου διάθεσης. Η διάθεση

μπορεί να πραγματοποιηθεί είτε σε επίπεδο πλήρους ανάκτησης/συγκομιδής (harvesting), όπου πρακτικά μεταφορτώνεται το σύνολο (ή προκαθορισμένα υποσύνολα) των εγγραφών ενός αποθετηρίου 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).
 - Χρήση κοινών σχημάτων μεταδεδομένων (π.χ. Dublin Core, MODS, CDWA, EAD, κλπ.)
 - Κωδικοποίηση των τιμών των δεδομένων σύμφωνα με ένα κοινό πρότυπο. Για παράδειγμα, οι ημερομηνίες να αναγράφονται με τον ίδιο τρόπο σε όλα τα συστήματα.
- **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;). Το ιδανικό είναι η δήλωση της σημασίας κάθε στοιχείου να πραγματοποιείται μέσω μιας κατάλληλης γλώσσας αναπαράστασης γνώσης όπως RDF ή OWL και κατάλληλου εξειδικευμένου προτύπου όπως EDM (Europeana Data Model), CIDOC-CRM, LIDO, CERIF ή αντίστοιχου.



Εικόνα 2 Ψηφιακό περιεχόμενο, διαλειτουργικότητα και διατήρησή του.

Επιπλέον, για τη διαλειτουργικότητα σε σημασιολογικό επίπεδο είναι σημαντική απαίτηση η χρήση ελεγχόμενων λεξιλογίων όπου η τιμή ενός πεδίου μεταδεδομένων μπορεί να αντληθεί από έναν προκαθορισμένο κατάλογο τιμών (π.χ. λεξιλόγιο, κατάλογο καθιερωμένων όρων, θησαυρό). Συνήθη παραδείγματα αποτελούν πεδία όπως θεματική κατηγορία, γεωγραφική περιοχή, χρονολογική περίοδο, γλώσσα αλλά και αναφορά σε οντότητες όπως άτομα και οργανισμούς. Ιδανικά, κάθε αναφορά στα μεταδεδομένα εγγραφών σε τιμή ελεγχόμενου πεδίου θα πρέπει να περιέχει μοναδικό αναγνωριστικό της τιμής σε καθιερωμένο ελεγχόμενο λεξιλόγιο / θησαυρό. Το καθιερωμένο λεξιλόγιο μπορεί να οριστεί και από τον φορέα που πραγματοποιεί την τεκμηρίωση αν δεν υπάρχει αντίστοιχο διαθέσιμο λεξιλόγιο/θησαυρός ευρείας χρήσης.

Η επίτευξη της ιδανικής διαλειτουργικότητας σε όλα τα επίπεδα αποτελεί εξαιρετικά δύσκολο εγχείρημα, ιδιαίτερα όταν αφορά μεγάλο πλήθος αναγκαστικά ετερογενούς ανοικτού ψηφιακού περιεχομένου. Προτείνεται μια **συνολική, ρεαλιστική προσέγγιση που θα ορίζει διαφορετικά επίπεδα απαιτούμενων χαρακτηριστικών** που θα υποστηρίζουν αντίστοιχα επίπεδα επιθυμητών υπηρεσιών ανάλογα με τον τύπο, το πλήθος και τις ιδιαιτερότητες των επιμέρους κατηγοριών συστημάτων. Για παράδειγμα, οι γενικοί κανόνες που θα αφορούν όλα τα συστήματα θα είναι πιο χαλαροί από ό,τι οι αντίστοιχοι για μια συγκεκριμένη, οριοθετημένη κατηγορία συστημάτων (π.χ. ψηφιακές βιβλιοθήκες, μουσεία, αρχεία), όπου μπορεί να απαιτείται συμβατότητα με πιο αναλυτικά και εξειδικευμένα σχήματα μεταδεδομένων.

Τα προτεινόμενα επίπεδα διαλειτουργικότητας περιγράφονται αναλυτικά στον Πίνακα 3. Τα βασικά τους σημεία μπορούν να συνοψιστούν ως εξής:

- **Πρώτο επίπεδο:** Εξασφαλίζει online διάθεση των μεταδεδομένων ως Ανοικτών Δημόσιων Δεδομένων (open data) και πρόσβαση στο περιεχόμενο με μόνιμους προσδιοριστές. Δίνει τη δυνατότητα για online έλεγχο του περιεχομένου του κάθε αποθετηρίου ανά πάσα στιγμή (π.χ. στοιχεία για το πόσες εγγραφές και τεκμήρια υπάρχουν). Δίνει τη δυνατότητα για ανάκτηση μεταδεδομένων που σχετίζονται με θέματα πνευματικής ιδιοκτησίας και κανόνες διάθεσης του περιεχομένου (το σημείο αυτό είναι ακόμα υπό διαμόρφωση και διεθνώς, οπότε θα πρέπει στα επόμενα χρόνια να υπάρχει πρόβλεψη για σταδιακή συμμόρφωση στα πρότυπα που θα έχουν διαμορφωθεί). Αντιστοιχεί στη διασφάλιση διαλειτουργικότητας επιπέδου συστημάτων αποθετηρίων με κάποια βασικά στοιχεία διαλειτουργικότητας σύνταξης και δομής.
- **Δεύτερο επίπεδο:** Εξασφαλίζει online διάθεση ευρύτερου συνόλου μεταδεδομένων ως Ανοικτά Δημόσια Δεδομένα σε επίπεδο ιδιαίτερα κατάλληλο για ανάπτυξη εφαρμογών από τρίτους. Δίνει τη δυνατότητα για βασικές υπηρεσίες ενιαίας πλοήγησης και οπτικοποίησης του περιεχομένου. Εξασφαλίζει ένα βασικό επίπεδο συμβατότητας με το σημασιολογικό ιστό. Αντιστοιχεί στη διασφάλιση διαλειτουργικότητας σύνταξης και δομής.
- **Τρίτο επίπεδο:** Εξασφαλίζει online διάθεση αναλυτικών μεταδεδομένων ως Ανοικτών Δημόσιων Δεδομένων σε επίπεδο που επιτρέπει την ανάπτυξη πληθώρας υπηρεσιών προστιθέμενης αξίας από τρίτους. Δίνει τη δυνατότητα για υψηλού επιπέδου υπηρεσίες ενιαίας αναζήτησης, πλοήγησης και οπτικοποίησης του περιεχομένου. Εξασφαλίζει διαλειτουργικότητα σε επίπεδο ελεγχόμενων λεξιλογίων. Εξασφαλίζει ένα ικανοποιητικό επίπεδο σημασιολογικής συμβατότητας. Αντιστοιχεί στη διασφάλιση σημασιολογικής διαλειτουργικότητας.

Επίπεδο	Περιγραφή	Υποστηριζόμενες υπηρεσίες
1	Διάθεση μέσω web services για harvesting των μεταδεδομένων, κατ'ελάχιστον των 15 πεδίων του βασικού σχήματος Dublin Core (http://purl.org/dc/elements/1.1) όπου έχουν εφαρμογή. Υποστήριξη μοναδικών προσδιοριστών ανά εγγραφή μεταδεδομένων. Ευρητηρίαση τουλάχιστον των μεταδεδομένων από τις κύριες διαδικτυακές μηχανές αναζήτησης γενικού σκοπού. Διάθεση πληροφορίας για την κατάσταση του περιεχομένου όσον αφορά θέματα πνευματικής ιδιοκτησίας.	<ol style="list-style-type: none"> 1. Ανάκτηση βασικών μεταδεδομένων για χρήση σε άλλες εφαρμογές. 2. Παραγωγή αναφορών για το περιεχόμενο του συνόλου των συστημάτων. 3. Ενιαία αναζήτηση σε βασικά πεδία.
2	Διάθεση μέσω web services για harvesting των μεταδεδομένων και μετα-αναζήτηση. Παροχή μεταδεδομένων σε μορφή Europeana Semantic Elements. Προσδιορισμός συγκεκριμένων sets για επιλεκτικό harvesting. Κωδικοποίηση τιμών πεδίων σύμφωνα με διεθνή πρότυπα.	<ol style="list-style-type: none"> 1. Ανάκτηση του συνόλου των μεταδεδομένων για χρήση σε άλλες εφαρμογές. 2. Παραγωγή αναφορών για το περιεχόμενο του συνόλου των συστημάτων. 3. Ενιαία αναζήτηση σε βασικά πεδία. 4. Βασικές λειτουργίες ενιαίας πλοήγησης και οπτικοποίησης.
3	Υποστήριξη διεθνώς καθιερωμένου σχήματος μεταδεδομένων κατάλληλου για ακριβή σημασιολογική αναπαράσταση. Χρήση στα μεταδεδομένα καθιερωμένων λεξιλογίων ευρείας χρήσης (π.χ. θησαυροί, θεματικές επικεφαλίδες, γεωγραφική πληροφορία). Διάθεση μεταδεδομένων ως linked data.	<ol style="list-style-type: none"> 1. Ανάκτηση του συνόλου των μεταδεδομένων για χρήση σε άλλες εφαρμογές. 2. Παραγωγή αναφορών για το περιεχόμενο του συνόλου των συστημάτων. 3. Ενιαία αναζήτηση σε βασικά πεδία. 4. Εξελιγμένες λειτουργίες ενιαίας πλοήγησης και οπτικοποίησης. 5. Διασύνδεση με άλλα σύνολα μεταδεδομένων (linked data).

Πίνακας 3. Επίπεδα διαλειτουργικότητας αποθετηρίων περιεχομένου.

Η συγκεκριμένη προσέγγιση είναι σύμφωνη και με την πρόσφατη διεθνή εμπειρία σε προσπάθειες συγκέντρωσης και διάθεσης περιεχομένου ιδιαίτερα μεγάλης κλίμακας, όπως οι ευρωπαϊκές πύλες Europeana ^[10] και DRIVER ^[11]. Στις αντίστοιχες περιπτώσεις, το μοντέλο που υιοθετήθηκε ήταν ο προσδιορισμός συγκεκριμένων ελάχιστων προϋποθέσεων που αφορούν το σύνολο των αποθετηρίων, ενώ όταν είναι δυνατή η κατηγοριοποίηση των επιμέρους αποθετηρίων προβλέπονται πιο λεπτομερείς προδιαγραφές και πρότυπα (π.χ. το πρότυπο LIDO για περιεχόμενο μουσείων και πολιτιστικής κληρονομιάς ^{[13][14][15]}). Παράλληλα, προβλέπεται και η κωδικοποίηση των λεπτομερών (ανά κατηγορία υλικού) προτύπων μεταδεδομένων σύμφωνα με το Europeana Data Model για κατάλληλες κατηγορίες περιεχομένου (π.χ. πολιτιστικό περιεχόμενο).

Τέλος, επισημαίνεται ότι οι προτεινόμενες προδιαγραφές, και ιδιαίτερα αυτές που αφορούν συγκεκριμένα πεδία μεταδεδομένων, αφορούν μόνο τη διαλειτουργικότητα (π.χ. την ανάκτηση των μεταδεδομένων μέσω κατάλληλων πρωτοκόλλων) και δεν είναι κατ' ανάγκην περιοριστικές για το βάσει ποιων προτύπων γίνεται η πρωτογενής καταλογογράφηση στο κάθε αποθετήριο. Η προσέγγιση αυτή είναι σύμφωνη με τις αρχές των shareable metadata, όπου διευκρινίζεται σαφώς πως τα μεταδεδομένα που καταγράφει και διατηρεί ένας οργανισμός για το ψηφιακό του υλικό μπορεί να είναι πολύ διαφορετικά από αυτά που διαθέτει δημόσια μέσω web services ^[12].

Παρακάτω παρατίθενται ενδεικτικές προδιαγραφές διαλειτουργικότητας σε μορφή απαιτήσεων πινάκων συμμόρφωσης για κάθε επίπεδο διαλειτουργικότητας.

Προδιαγραφές διαλειτουργικότητας αποθετηρίων

Α/Α	Απαίτηση	Προδιαγραφή
1.	Διαλειτουργικότητα σε επίπεδο συστημάτων	Για την κωδικοποίηση των μεταδεδομένων χρησιμοποιείται το πρότυπο UTF-8.
2.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής	Τα μεταδεδομένα κάθε εγγραφής είναι διαθέσιμα για ανάκτηση (harvesting) σε μορφή Unqualified Dublin Core (ISO 15836:2003).
3.	Διαλειτουργικότητα σε επίπεδο συστημάτων	Θα πρέπει να υποστηρίζεται μόνιμη πρόσβαση σε ανάκτηση των μεταδεδομένων βάσει του πρωτοκόλλου OAI-PMH, έκδοση 2.0 με υποστήριξη όλων των verbs του πρωτοκόλλου.
4.	Διαλειτουργικότητα σε επίπεδο συστημάτων	Θα πρέπει να υποστηρίζεται η διάθεση των μεταδεδομένων μέσω τουλάχιστον ενός πρωτοκόλλου διάθεσης / μετα-αναζήτησης, συγκεκριμένα ένα εκ των SRU/SRW, Z39.50. Η συγκεκριμένη λειτουργία θα πρέπει να διατίθεται δημόσια, χωρίς περιορισμούς πρόσβασης.
5.	Διαλειτουργικότητα σε επίπεδο συστημάτων	Πρέπει να υποστηρίζεται η ευρετηρίαση τουλάχιστον των μεταδεδομένων από συνήθεις διαδικτυακές μηχανές αναζήτησης (Google, Bing, Yahoo).
6.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής	Τα μεταδεδομένα κάθε εγγραφής είναι διαθέσιμα για ανάκτηση (harvesting) σε μορφή Dublin Core (ISO 15836:2009). Τα δεκαπέντε (15) βασικά πεδία του Dublin Core namespace dc (http://purl.org/dc/elements/1.1) θα πρέπει να είναι υποχρεωτικά για κάθε εγγραφή, όπου έχουν εφαρμογή.
7.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής. Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Τα μεταδεδομένα κάθε εγγραφής είναι μόνιμως διαθέσιμα για ανάκτηση (harvesting) σε μορφή Europeana Semantic Elements (έκδοση 3.4.1). Η αντιστοίχιση του πλήρους σχήματος μεταδεδομένων στο πρότυπο Europeana Semantic Elements (ESE) είναι μόνιμως διαθέσιμα για ανάκτηση. Θα πρέπει να παρέχεται ανά εγγραφή το σύνολο των πεδίων μεταδεδομένων που μπορούν να αναπαρασταθούν σε μορφή Europeana Semantic Elements (έκδοση 3.4.1) – όχι μόνο αυτά που προορίζονται κατά περίπτωση για διοχέτευση στη Europeana.
8.	Διαλειτουργικότητα σε επίπεδο συστημάτων. Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής.	Θα πρέπει να ορίζονται με τη μορφή OAI-PMH Sets και γίνονται διαθέσιμα για ανάκτηση μέσω OAI-PMH ω2.0 σε μορφή Dublin Core υποσύνολα των εγγραφών του αποθετηρίου, κατ' ελάχιστον τα ακόλουθα: i. Το σύνολο των εγγραφών που περιέχουν πλήρες κείμενο. ii. Από ένα σύνολο εγγραφών για κάθε διαφορετική τιμή του πεδίου dc.type.

A/A	Απαίτηση	Προδιαγραφή
9.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής.	Στα Dublin Core μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να περιέχεται ένας τουλάχιστον μόνιμος προσδιοριστής (persistent identifiers) που να ταυτοποιεί την εγγραφή στο πεδίο dc.identifier. Ο μόνιμος προσδιοριστής δεν επιτρέπεται οποτεδήποτε στο μέλλον να αλλάξει ή να αποδοθεί σε άλλο ψηφιακό πόρο. Ο μόνιμος προσδιοριστής θα πρέπει να ακολουθεί το διεθνές πρότυπο Handle.
10.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής.	Στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH προς τον ενιαίο κατάλογο θα πρέπει για κάθε ένα από τα ψηφιακά αρχεία που συνδέονται με την εγγραφή να περιέχεται λειτουργούσα διεύθυνση (URL) η οποία να επιτρέπει την ανάκτηση υλικού κατάλληλου για την προβολή/προεπισκόπηση του κάθε αρχείου. Παραδείγματα: εικόνα ανάλυσης επιπέδου thumbnail για εικόνες, εικόνα εξωφύλλου για βιβλία, εικόνα πρώτης σελίδας για άρθρα, κώδικας ενσωμάτωσης σε τρίτη ιστοσελίδα για streaming video.
11.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής. Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Η κωδικοποίηση του πεδίου dc.type στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να περιέχει μοναδικό αναγνωριστικό με αναφορά σε τυποποιημένο λεξιλόγιο όρων για τύπους. Το τυποποιημένο λεξιλόγιο μπορεί να οριστεί και από τον φορέα που κάνει την τεκμηρίωση αν δεν υπάρχει αν δεν υπάρχει αντίστοιχο διαθέσιμο λεξιλόγιο/θησαυρός ευρείας χρήσης.
12.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής	Η κωδικοποίηση του πεδίου dc.creator στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να ακολουθεί καθιερωμένο πρότυπο βιβλιογραφικών αναφορών σε ό,τι αφορά τη γραφή των ονομάτων δημιουργών.
13.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής	Η κωδικοποίηση του πεδίου dc.language (αν υπάρχει) στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να ακολουθεί το πρότυπο ISO 639-2, ενώ για τις περιπτώσεις γλωσσών που στο συγκεκριμένο πρότυπο έχουν δύο διαφορετικούς κωδικούς (ορολογίας και βιβλιογραφικό), θα πρέπει να χρησιμοποιείται ο βιβλιογραφικός κώδικας (ISO 639-2/B).
14.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής	Η κωδικοποίηση του πεδίων ημερομηνίας (dc.date & qualifiers) στα μεταδεδομένα εγγραφής που διατίθεται μέσω OAI-PMH θα πρέπει να ακολουθεί το πρότυπο ISO 8601.
15.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής	Στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH το πεδίο dc.date θα αναφέρεται στην ημερομηνία έκδοσης του τεκμηρίου, όπου υπάρχει έκδοση. Διαφορετικά θα συμπληρώνεται η ημερομηνία δημιουργίας του τεκμηρίου στο πεδίο dcterms.created (http://purl.org/dc/terms/created).

A/A	Απαίτηση	Προδιαγραφή
16.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής μεταδεδομένων Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Πρωτότυπα δεσμευμένα λεξιλόγια, θησαυροί και συναφή εργαλεία ορολογίας που αναπτύσσονται στο πλαίσιο της τεκμηρίωσης του αποθετηρίου θα πρέπει να διατίθενται για εξαγωγή σε μορφή συμβατή με ένα από τα παρακάτω πρότυπα: Simple Knowledge Organization System (SKOS), ISO 2788, ISO 5964, ISO 25964-1
17.	Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Κάθε αναφορά στα μεταδεδομένα εγγραφών σε θεματικές κατηγορίες θα πρέπει να περιέχει μοναδικό αναγνωριστικό τους σε καθιερωμένο ελεγχόμενο λεξιλόγιο ή/και θησαυρό.
18.	Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Κάθε αναφορά στα μεταδεδομένα εγγραφών σε ονόματα ανθρώπων θα πρέπει να περιέχει μοναδικό αναγνωριστικό τους σε καθιερωμένο ελεγχόμενο λεξιλόγιο / θησαυρό. Το τυποποιημένο λεξιλόγιο μπορεί να οριστεί και από τον φορέα που κάνει αν δεν υπάρχει αν δεν υπάρχει αντίστοιχο διαθέσιμο λεξιλόγιο/θησαυρός ευρείας χρήσης.
19.	Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Κάθε αναφορά στα μεταδεδομένα εγγραφών σε τοπωνύμια θα πρέπει να περιέχει μοναδικό αναγνωριστικό τους σε καθιερωμένο ελεγχόμενο λεξιλόγιο (π.χ. geonames) ή/και θησαυρό.
20.	Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Κάθε αναφορά στα μεταδεδομένα εγγραφών σε ονόματα οργανισμών θα πρέπει να περιέχει μοναδικό αναγνωριστικό τους σε καθιερωμένο ελεγχόμενο λεξιλόγιο ή/και θησαυρό. Το τυποποιημένο λεξιλόγιο μπορεί να οριστεί και από τον φορέα που κάνει την τεκμηρίωση αν δεν υπάρχει αν δεν υπάρχει αντίστοιχο διαθέσιμο λεξιλόγιο/θησαυρός ευρείας χρήσης.
21.	Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Κάθε αναφορά στα μεταδεδομένα εγγραφών σε χρονολογικές περιόδους θα πρέπει να περιέχει μοναδικό αναγνωριστικό τους σε καθιερωμένο ελεγχόμενο λεξιλόγιο ή/και θησαυρό. Το τυποποιημένο λεξιλόγιο μπορεί να οριστεί και από τον φορέα που κάνει την τεκμηρίωση αν δεν υπάρχει αν δεν υπάρχει αντίστοιχο διαθέσιμο λεξιλόγιο/θησαυρός ευρείας χρήσης.
22.	Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Το αποθετήριο χρησιμοποιεί για την τεκμηρίωση κάποιο από τα διεθνώς αποδεκτά καθιερωμένα σχήματα και οντολογίες για την τεκμηρίωση των Ψηφιακών του Πόρων με δυνατότητες αναλυτικής καταγραφής της σημασιολογίας των μεταδεδομένων, όπως MARC21, UNIMARC, MODS, EAD, Europeana Data Model, CERIF, LIDO, VRA Core, CIDOC-CRM.
23.	Διαλειτουργικότητα σε επίπεδο συστημάτων. Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Τα μεταδεδομένα του αποθετηρίου είναι διαθέσιμα ως Linked Data.

Πίνακας 4. Προδιαγραφές για όλες τις κατηγορίες αποθετηρίων (Επίπεδο 1)

Για αντίστοιχα επίπεδα διαλειτουργικότητας σε διαφορετικά είδη αποθετηρίων, (με βάση το κύριο περιεχόμενο), οι προδιαγραφές θα επικαιροποιηθούν αντίστοιχα.

3.3. Μόνιμοι προσδιοριστές (διευθύνσεις) πόρων ανοικτού ψηφιακού περιεχομένου

Απαραίτητη για την ύπαρξη διαρκούς διαλειτουργικότητας είναι η σταθερή διευθυνσιοδότηση του κάθε ψηφιακού πόρου, ανεξάρτητα από το συγκεκριμένο σύστημα λογισμικού και τη διαδικτυακή διεύθυνση που χρησιμοποιείται για τη διάθεσή του, ώστε να εξασφαλίζεται η πρόσβαση σε αυτόν ανεξάρτητα από το φορέα και το σύστημα λογισμικού που τον παρέχει. Απαιτείται παροχή των τεκμηρίων με αποθετήριο που να υποστηρίζει μηχανισμό persistent identifier handle service για την παροχή μόνιμων τοποθεσιών ψηφιακού περιεχομένου **ανεξάρτητους από το συγκεκριμένο σύστημα λογισμικού που υλοποιεί το αποθετήριο και τη διαδικτυακή διεύθυνση στην οποία παρέχεται**. Το σύστημα για την παροχή αυτού του μηχανισμού είναι το πρότυπο Handle System RFC3650 ή/και το σύστημα DOI. Σε αντίθεση με το σύστημα DOI στο σύστημα Handle δεν υπάρχει κόστος λήψης μόνιμης διεύθυνσης ανά ψηφιακό πόρο.

Προδιαγραφές για μόνιμους προσδιοριστές

A/A	Απαίτηση	Προδιαγραφή
1.	Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier)	Παροχή των τεκμηρίων με αποθετήριο που να υποστηρίζει μηχανισμό persistent identifier.
2.	Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier)	Κάθε μόνιμος προσδιοριστής θα πρέπει να είναι ανεξάρτητος από το διαδικτυακό domain διάθεσης του αποθετηρίου, το φορέα διάθεσης και το λογισμικό του αποθετηρίου.
3.	Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier)	Έκδοση από κατάλληλη αρχή και ανάθεση ενός μόνιμου προσδιοριστή (persistent identifier) σύμφωνα με το πρότυπο Handle (RFC3650, RFC3652) σε κάθε εγγραφή μεταδεδομένων στο αποθετήριο.
4.	Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier)	Το λογισμικό διάθεσης να υποστηρίζει την μόνιμη διάθεση όλων των εγγραφών μεταδεδομένων και του ψηφιακού περιεχομένου στην διεύθυνση που ορίζει ο μόνιμος προσδιοριστής τους.
5.	Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier)	Το λογισμικό θα πρέπει να υποστηρίζει Handle System RFC3650.
6.	Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier)	Ο κάθε μόνιμος προσδιοριστής δεν επιτρέπεται οποτεδήποτε στο μέλλον να αλλάξει ή να αποδοθεί σε άλλο ψηφιακό πόρο

Πίνακας 5 Προδιαγραφές για απαίτηση μόνιμων προσδιοριστών ψηφιακών πόρων

3.4. Προδιαγραφές για διάθεση ανοικτού ψηφιακού περιεχομένου

Η παρουσίαση και διάθεση περιεχομένου σε τελικούς χρήστες αποτελεί ιδιαίτερα κρίσιμη παράμετρο στο σύγχρονο περιβάλλον. Μία απαραίτητη και βασική προϋπόθεση για την αποτελεσματική παρουσίαση περιεχομένου σε τελικούς χρήστες είναι η συμμόρφωση με τις προδιαγραφές του Πλαισίου Πιστοποίησης Δημόσιων Διαδικτυακών Τόπων [16].

Ωστόσο, για την επίτευξη των στόχων του προγράμματος απαιτείται η διαμόρφωση μιας ελκυστικής «**εμπειρίας χρήστη**» που είναι απαραίτητη για την ευρύτερη δυνατή διάχυση κάθε είδους περιεχομένου στο ενδιαφερόμενο κοινό αλλά και για την επαναχρησιμοποίηση της, με βάση τις πλήρης δυνατότητες της.

Είναι απαραίτητο να σημειωθεί πως οι σχετικές προδιαγραφές είναι σε μεγάλο βαθμό συναρτώμενες με τον τύπο του ψηφιακού πόρου (π.χ. κείμενο, εικόνα, οπτικοακουστικό). Συνεπώς, οι προδιαγραφές διακρίνονται σε δύο κατηγορίες, (α) ανεξάρτητες του τύπου του υλικού, (β) εξειδικευμένες σε συγκεκριμένο τύπο ψηφιακού πόρου.

Παρακάτω παρατίθενται αναλυτικοί πίνακες με τις προδιαγραφές της κατηγορίας (α) (σε δύο επίπεδα) καθώς και ενδεικτικές προδιαγραφές για την κατηγορία (β) και ειδικότερα για τους τύπους υλικού κείμενο και εικόνα. Οι προδιαγραφές αυτές έχουν προκύψει με βάση την εμπειρία από τα υπάρχοντα συστήματα διάθεσης περιεχομένου καθώς και από ελληνικά αλλά και διεθνή καλά παραδείγματα υλοποίησης.

Υποχρεωτικές προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού (Επίπεδο 1)

A/A	Απαίτηση	Προδιαγραφή
1.	Προβολή του υλικού σε κατάλληλο context.	Για κάθε τεκμήριο υπάρχει διαδικτυακή σελίδα παρουσίασης που περιλαμβάνει τα μεταδεδομένα του, ενώ γίνεται προβολή του ίδιου του ψηφιακού τεκμηρίου. Σε κάποιες περιπτώσεις στη σελίδα μπορεί να περιλαμβάνεται μόνο υπερ-σύνδεσμος σε ξεχωριστή σελίδα προβολής του τεκμηρίου.
2.	Εύκολος εντοπισμός και ανάκτηση του υλικού από το διαδικτυακό χρήστη.	Δυνατότητα download των τεκμηρίων από τον τελικό χρήστη (όπου είναι εφικτό βάσει πνευματικών δικαιωμάτων) μέσω εμφανούς επιλογής / υπερ-συνδέσμου στη σελίδα παρουσίασης του κάθε τεκμηρίου.
3.	Προβολή του υλικού σε κατάλληλο context. Διασύνδεση του υλικού από εξωτερικές πηγές.	Οι σελίδες παρουσίασης του τεκμηρίου καθώς και οι ξεχωριστές σελίδες προβολής του (όπου υπάρχουν) είναι απευθείας προσπελάσιμες (deep linking) μέσω φιλικών προς το χρήστη διαδικτυακών διευθύνσεων (URL). Οι διευθύνσεις των σελίδων παρουσίασης του τεκμηρίου διαθέτουν μοναδικούς, μόνιμους προσδιοριστές (persistent identifiers).
4.	Εύκολος εντοπισμός και ανάκτηση του υλικού από το διαδικτυακό χρήστη.	Απλή και σύνθετη αναζήτηση στα μεταδεδομένα τεκμηρίων. Στον κατάλογο αποτελεσμάτων περιλαμβάνονται υπερ-συνδέσμοι στις σελίδες παρουσίασης των αντίστοιχων τεκμηρίων.

Πίνακας 6. Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού (Επίπεδο 1)

**Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού
(Επίπεδο 2)**

A/A	Απαίτηση	Προδιαγραφή
1.	Πολυκαναλική διάθεση	Διαθεσιμότητα του περιεχομένου – με κατάλληλες προσαρμογές – για προβολή σε συσκευές εναλλακτικές των συνηθών desktop / laptop υπολογιστών όπως e-book readers, net-books, tablets, smart phones.

**Πίνακας 7. Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού
(Επίπεδο 2)**

Υποχρεωτικές προδιαγραφές παρουσίασης και διάθεσης για κείμενα

A/A	Απαίτηση	Προδιαγραφή
1.	Βελτίωση εμπειρίας χρήστη.	Παρουσίαση του κειμένου ως ενιαίου αρχείου – δεν αρκεί παράθεση υπερ-συνδέσμων σε επιμέρους σαρωμένες σελίδες, προσβάσιμες με επιπλέον βήματα.
2.	Βελτίωση εμπειρίας χρήστη.	Δυνατότητα online ανάγνωσης του κειμένου μέσω κατάλληλης εφαρμογής φυλλομέτρησης / παρουσίασης ως e-book. Θα πρέπει να περιλαμβάνονται δυνατότητες μεγέθυνσης σε επίπεδο που η online ανάγνωση να είναι άνετη για τους διαδικτυακούς χρήστες ενώ δεν θα πρέπει να προτιμάται η υλοποίηση της λειτουργίας μέσω κλειστών εξωτερικών, από τον browser, συστημάτων.
3.	Εύκολος εντοπισμός και ανάκτηση υλικού από το διαδικτυακό χρήστη.	Αναζήτηση στο πλήρες κείμενο. Εξαιρούνται περιπτώσεις τεκμηρίων που προέρχονται από σάρωση και δεν είναι εφικτό να πραγματοποιηθεί οπτική αναγνώριση χαρακτήρων (OCR) με υψηλά ποσοστά επιτυχίας.

Πίνακας 8. Προδιαγραφές παρουσίασης και διάθεσης για κείμενα

Υποχρεωτικές προδιαγραφές παρουσίασης και διάθεσης για εικόνες

A/A	Απαίτηση	Προδιαγραφή
1.	Βελτίωση εμπειρίας χρήστη.	Μεγέθυνση / σμίκρυνση πολλών επιπέδων με ταυτόχρονη δυνατότητα "pan", όπου είναι εφικτό βάσει πνευματικών δικαιωμάτων. Συνιστάται η ταυτόχρονη παρουσίαση στην ίδια σελίδα τόσο του μεγεθυμένου τμήματος της εικόνας όσο και ολόκληρης της εικόνας.
2.	Βελτίωση εμπειρίας χρήστη.	Περιστροφή τουλάχιστον τριών επιπέδων κατά βήμα 90 μοιρών και προς τις δύο κατευθύνσεις, όπου κρίνεται χρήσιμο για τους διαδικτυακούς χρήστες.
3.	Βελτίωση εμπειρίας χρήστη.	Υπερ-σύνδεσμος για απευθείας προβολή της εικόνας στη μέγιστη ανάλυση που είναι διαθέσιμη στο Διαδίκτυο. Ο προσδιορισμός της μέγιστης ανάλυσης που διατίθεται διαδικτυακά μπορεί να εξαρτάται από περιορισμούς πνευματικών δικαιωμάτων ή από την καλύτερη δυνατή διαμόρφωση της εμπειρίας του χρήστη, π.χ. χρειάζεται πολλές φορές πρακτικά συμβιβασμός μεταξύ προβολής σε υψηλή ανάλυση και διατήρησης του μεγέθους αρχείου σε λογικά επίπεδα.
4.	Βελτίωση εμπειρίας χρήστη.	Η υποστήριξη των προδιαγραφών 1, 2, 3 θα πρέπει να εισάγει τις μικρότερες δυνατές απαιτήσεις από τα λογισμικά πλοήγησης (browsers) διαδικτυακών χρηστών. Κατά προτίμηση θα πρέπει να υπάρχει έκδοση της παρουσίασης στην οποία απαιτείται μόνο υποστήριξη Javascript. Είναι επίσης αποδεκτό, αλλά δεν συνιστάται, οι λειτουργίες να βασίζονται αποκλειστικά σε τεχνολογίες Flash, Java, Silverlight ή αντίστοιχες. Μια βέλτιστη εναλλακτική είναι η διαθεσιμότητα τουλάχιστον δύο εκδόσεων, μία μικρών απαιτήσεων (javascript) και μία πιο απαιτητική (π.χ. Flash, Java ή Silverlight).
5.	Βελτίωση εμπειρίας χρήστη.	Συνιστάται οι προ-αναφερθείσες λειτουργίες να διατίθενται μέσα από τη σελίδα παρουσίασης του τεκμηρίου (με μεταδεδομένα και προβολή της εικόνας) ώστε να μη χρειάζεται αλλαγή σελίδας από το διαδικτυακό χρήστη.

Πίνακας 9. Προδιαγραφές παρουσίασης και διάθεσης για εικόνες

3.5. Διαλειτουργικότητα με καταλόγους και e-υποδομή ανοικτού ψηφιακού περιεχομένου ΕΚΤ

Για την αύξηση της χρήσης του περιεχομένου είναι επιθυμητή η διαλειτουργικότητα με α) συλλογικό κατάλογο (μητρώο ανοικτού ψηφιακού περιεχομένου) και β) το σύστημα ασφαλούς διαφύλαξης ανοικτού ψηφιακού περιεχομένου, με συνολικό σκοπό την δημιουργία ενιαία αναζητήσιμου ψηφιακού αποθέματος, καθώς και για την παράλληλη εφαρμογή βασικών μηχανισμών ασφάλειας και ψηφιακής διατήρησης του ανοικτού ψηφιακού περιεχομένου που έχει παραχθεί.

Η e-υποδομή (μητρώο ανοικτού ψηφιακού τεκμηριωμένου περιεχομένου) μπορεί να συλλέγει – με περιοδική αυτόματη ανανέωση σε τακτά χρονικά διαστήματα – τα

μεταδεδομένα τεκμηρίωσης των επιμέρους αποθετηρίων σε σύστημα με σκοπό την δημιουργία ενιαίου καταλόγου. Σημειώνεται ότι οι παρεχόμενες προδιαγραφές αφορούν διαλειτουργικότητα με οποιαδήποτε σύστημα συσσωρευσης που λειτουργεί με βάση τα αντίστοιχα διεθνώς καθιερωμένα ανοικτά πρότυπα. Επιπρόσθετα, για λόγους βασικής ασφάλειας πληροφοριακών συστημάτων και προστασίας της επένδυσης σε έργα ψηφιοποίησης δίνεται η δυνατότητα για διαλειτουργικότητα με σύστημα ασφαλούς διαφύλαξης περιεχομένου, το οποίο παρέχεται από το ΕΚΤ και αφορά την αυτοματοποιημένη συλλογή αντιγράφων, για λόγους ασφαλείας, του ψηφιακού περιεχομένου στην e-υποδομή του Εθνικού Κέντρου Τεκμηρίωσης, για όποιον φορέα το επιθυμεί. Το σύστημα αυτό αποτελεί προαπαιτούμενο για την μελλοντική εφαρμογή βασικών πολιτικών διατήρησης και στο ίδιο το ψηφιακό υλικό.

Έτσι διασφαλίζεται ένα τουλάχιστον **απομακρυσμένο αντίγραφο** των δεδομένων, μεταδεδομένων και του πρωτότυπου ψηφιακού περιεχομένου λειτουργώντας ως **remote data replication/disaster recovery** σύστημα στο τομέα του ψηφιακού περιεχομένου, αξιοποιώντας ήδη διαθέσιμα χαρακτηριστικά και διεπαφές των αποθετηρίων (οργανωμένα μεταδεδομένα, διεπαφές, μοναδικοί αναγνωριστές) για την επίτευξη βασικών όρων ασφαλείας και διαθεσιμότητας του περιεχομένου.

A/A	Απαίτηση	Περιγραφή
1.	Εναπόθεση μεταδεδομένων αποθετηρίων σε μητρώο ανοικτού ψηφιακού περιεχομένου	Παροχή μηχανισμών ανταλλαγής δεδομένων και μεταδεδομένων με το πρότυπο: <ul style="list-style-type: none"> ▪ OAI-PMH
2.	Συλλογή ψηφιακών πόρων αποθετηρίων σε υποδομή ασφαλούς διαφύλαξης ανοικτού ψηφιακού περιεχομένου	METS ή OAI-ORE μηχανισμοί ή άλλοι τεκμηριωμένα αντίστοιχοι για μεταφόρτωση αντιγράφων ασφαλείας σε κεντρική υποδομή
3.	Συλλογή ψηφιακών πόρων αποθετηρίων σε υποδομή ασφαλούς διαφύλαξης ανοικτού ψηφιακού περιεχομένου	Web service με δυνατότητα authentication/access restriction για μεταφόρτωση ψηφιακού αντικειμένων με βάση την μόνιμη (persistent) διεύθυνση του αντίστοιχου ψηφιακού πόρου/τεκμηρίου

Πίνακας 10. Διαλειτουργικότητα με μητρώο ανοικτού ψηφιακού περιεχομένου και σύστημα ασφαλούς διαφύλαξης ανοικτού ψηφιακού περιεχομένου

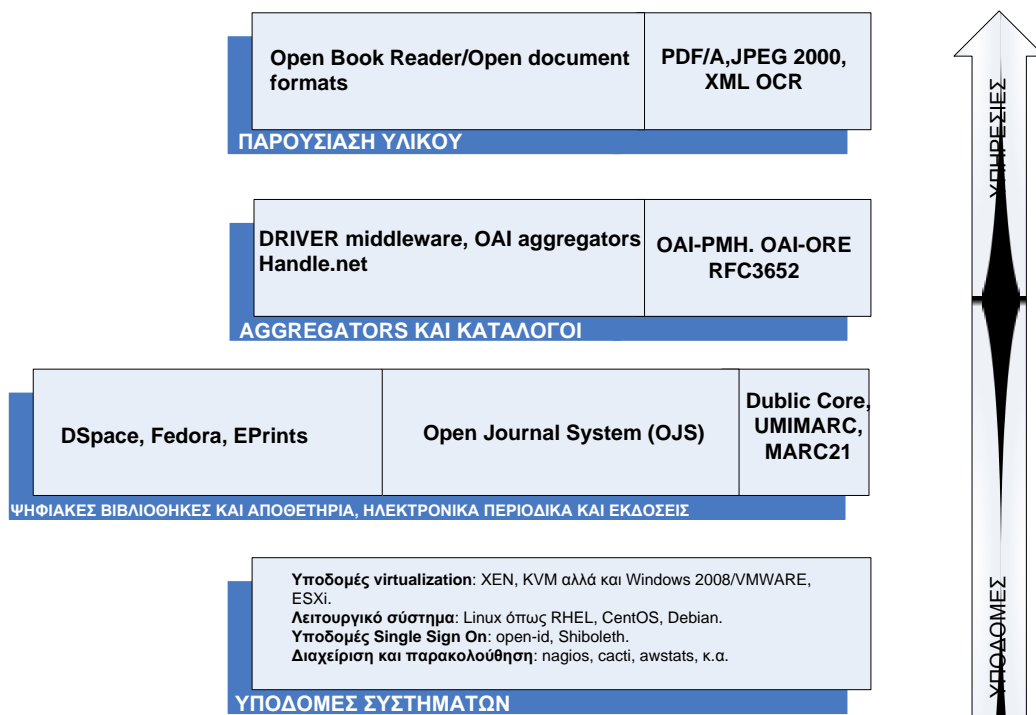
4. Ενδεικτικό λογισμικό για την ανάπτυξη υποδομών διάθεσης ανοικτού ψηφιακού περιεχομένου

Στον Πίνακα 13 αναφέρονται ενδεικτικά συστήματα τα οποία μπορούν να αξιοποιηθούν για την πλήρη υλοποίηση έργων αποθετηρίων με βάση τις παραπάνω προδιαγραφές. Αναφέρονται συστήματα που είναι διαθέσιμα μόνο ως ΕΛ/ΛΑΚ χωρίς να σημαίνει ότι δεν υπάρχουν και επιπλέον αντίστοιχα συστήματα ανοικτού ή και κλειστού λογισμικού. Σκοπός είναι να αναδειχθεί το εξειδικευμένο λογισμικό για συστήματα ψηφιακών βιβλιοθηκών/αποθετηρίων/εκδόσεων τα οποία έχουν υιοθετήσει πρότυπα που βρίσκονται υπό ανάπτυξη και τις τελευταίες τεχνολογικές εξελίξεις στους αντίστοιχους τομείς.

Τα συστήματα αυτά είναι πλήρως διαλειτουργικά και υλοποιούν μια πλήρη «στοίβα» (Εικόνα 3 Ενδεικτική «στοίβα» ΕΛ/ΛΑΚ για διάθεση ψηφιακού περιεχομένου σύμφωνα με ανοικτά πρότυπα) συστημάτων ελεύθερου και ανοικτού λογισμικού για την προσφορά υπηρεσιών ψηφιακών βιβλιοθηκών και αποθετηρίων με βάση διεθνή ανοικτά πρότυπα και με καλά κλιμακούμενο και προβλέψιμο κόστος και πολυπλοκότητα.

A/A	Περιοχή	Λογισμικό και πρότυπα
1.	Παρουσίαση ψηφιακών τεκμηρίων	Λογισμικά ΕΛ/ΛΑΚ <ul style="list-style-type: none"> Internet Archive Book Reader Flexpaper Multivio Πρότυπα <ul style="list-style-type: none"> PDF/A Jpeg hOCR ABBYY XML OCR PDF/A
2.	Aggregators και εθνικοί κατάλογοι	Συστήματα / λογισμικά ΕΛ/ΛΑΚ: <ul style="list-style-type: none"> DRIVER / D-NET Europeana Πρότυπα <ul style="list-style-type: none"> OAI-PMH OAI-ORE CERIF
3.	Handle services	<ul style="list-style-type: none"> HANDLE.NET RFC3652
4.	Διάθεση και οργάνωση περιεχομένου: Αποθετήρια Ηλεκτρονικά περιοδικά και εκδόσεις Ψηφιακές βιβλιοθήκες	Λογισμικά ΕΛ/ΛΑΚ: <ul style="list-style-type: none"> DSpace, Fedora E-prints Omeka Greenstone OJS openABEKT (υπό ανάπτυξη) Πρότυπα <ul style="list-style-type: none"> Dublic Core, MODS, UNIMARC, MARC 21, EDM, CDWA, LIDO, CIDOC CRM, CERIF OAI-PMH, Z39.50, SRU.W, OpenSearch, OpenURL
5.	Υποδομή συστημάτων: Λειτουργικά Συστήματα, Middleware/B.Δ. Πλατφόρμες virtualization και Cloud Συστήματα Single Sign On (SSON) Συστήματα διαχείρισης και παρακολούθησης	<ul style="list-style-type: none"> Linux Tomcat, JBoss, Postgress,MySQL XEN, KVM, G-Cloud ακόμα και Windows 2008/VMWARE, ESXi μιας και δρουν πολλαπλασιαστικά Shibboleth, openid nagios, cacti, graphite, puppet, awstats, κ.α.

Πίνακας 11. Ενδεικτική υλοποίηση συστημάτων διάθεσης ψηφιακού περιεχόμενου με ΕΛ/ΛΑΚ και σχετικά πρότυπα



Εικόνα 3 Ενδεικτική «στοίβα» ΕΛ/ΛΑΚ για διάθεση ψηφιακού περιεχομένου σύμφωνα με ανοικτά πρότυπα

5. Σύνοψη λειτουργικότητας χαρακτηριστικών και προδιαγραφών

Στον παρακάτω πίνακα γίνεται μια απόπειρα να αναδειχθεί πως οι επιμέρους προδιαγραφές που προτάθηκαν σε αυτή την μελέτη μπορούν να αποτελέσουν πράγματι προαπαιτούμενο για την επίτευξη κεντρικών στόχων των έργων και των επενδύσεων στην περιοχή της ανάδειξης του ελληνικού ψηφιακού περιεχομένου.

	Αύξηση διεθνούς χρήσης περιεχομένου (συμπερίληψη σε μηχανές αναζήτησης, full text indexing, κ.α.)	Αύξηση χρήσης του διατιθέμενου περιεχομένου στο Εθνικό επίπεδο	Επαναχρησιμοποίηση περιεχομένου	Δημιουργία υπηρεσιών προστιθέμενης αξίας	Εξασφάλιση της λειτουργίας, ασφάλειας, διατήρησης και επαναχρησιμοποίησης του περιεχομένου (Ψηφιακή διατήρηση)
Οπτική Αναγνώριση Χαρακτήρων (OCR) για υποστήριξη αναζήτησης και ευρετηρίασης του πλήρους κειμένου	√	√	√	√	√
Διαλειτουργικότητα με μητρώο ανοικτού τεκμηριωμένου περιεχομένου ΕΚΤ	√	√	√	√	√
Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής μεταδεδομένων και δεδομένων	√	√	√	√	√
Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	√	√	√	√	√
Διασύνδεση του υλικού από εξωτερικές πηγές.	√	√		√	√
Πολυκαναλική διάθεση	√	√		√	
Εύκολος εντοπισμός και ανάκτηση υλικού από το διαδικτυακό χρήστη.	√	√	√		
Βελτίωση εμπειρίας χρήστη. Προβολή του υλικού σε κατάλληλο context.	√	√			
Αποθήκευση δομημένου περιεχομένου σε κεντρικό τόπο για ψηφιακή διατήρηση/remote replication ψηφιακών τεκμηρίων			√		√

	Αύξηση διεθνούς χρήσης περιεχομένου (συμπερίληψη σε μηχανές αναζήτησης, full text indexing, κ.α.)	Αύξηση χρήσης του διατιθέμενου περιεχομένου στο Εθνικό επίπεδο	Επαναχρησιμοποίηση περιεχομένου	Δημιουργία υπηρεσιών προστιθέμενης αξίας	Εξασφάλιση της λειτουργίας, ασφάλειας, διατήρησης και επαναχρησιμοποίησης του περιεχομένου (Ψηφιακή διατήρηση)
Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier handle service)			√	√	√
Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier handle service)			√	√	√

Πίνακας 12. Σύνοψη απαιτούμενων λειτουργικών χαρακτηριστικών έργων περιεχομένου σε σχέση με επιδιωκόμενους στόχους

Βιβλιογραφία

- [1] «Οδηγός καλών πρακτικών για την ψηφιοποίηση και τη μακροπρόθεσμη διατήρηση του πολιτιστικού περιεχομένου», μελέτη στο πλαίσιο του Γ' ΚΠΣ, Επιχειρησιακό Πρόγραμμα "Κοινωνία της Πληροφορίας", Μέτρο 1.3, διαθέσιμο στην διεύθυνση <http://digitization.hpclab.ceid.upatras.gr/>, http://www.infosoc.gr/infosoc/el-GR/services/elibrary/reports_list/psifiopiisi/synoptikos_odigos.htm
- [2] Μελέτη ψηφιοποίησης τρισδιάστατων αντικειμένων, μελέτη στο πλαίσιο του Γ' ΚΠΣ, Επιχειρησιακό Πρόγραμμα "Κοινωνία της Πληροφορίας", Μέτρο 1.3, διαθέσιμο στην προσωρινή διεύθυνση http://www.infosoc.gr/infosoc/el-GR/services/elibrary/reports_list/psifiopiisi/3dimensions.htm
- [3] Κέντρο Πολιτισμικής Πληροφορικής, ΙΤΕ (2005, Απρίλιος). Οδηγός Πολιτισμικής Τεκμηρίωσης και Διαλειτουργικότητας, μελέτη στο πλαίσιο του Γ' ΚΠΣ, Επιχειρησιακό Πρόγραμμα "Κοινωνία της Πληροφορίας", Μέτρο 1.3, διαθέσιμο στη διεύθυνση <http://www.ics.forth.gr/CULTUREstandards>.
- [4] Cornell University Library, "Moving Theory into Practice: Digital Imaging Tutorial", διαθέσιμο στο <http://www.library.cornell.edu/preservation/tutorial/contents.html>
- [5] Harvard University Library, "Selection For Digitizing: A Decision-Making Matrix", <http://www.clir.org/pubs/reports/hazen/matrix.html>
- [6] International Federation of Library Associations and Institutions (IFLA), "Guidelines for digitization projects", διαθέσιμο στο <http://www.ifla.org/VII/s19/pubs/digit-guide.pdf>
- [7] National Archives and Records Administration (NARA), «NARA Guidelines for Digitizing Archival Materials for Electronic Access» http://www.archives.gov/research_room/arc/arc_info/guidelines_for_digitizing_archival_materials.html
- [8] Δημόσια διαβούλευση για τις γενικές αρχές και προτεραιότητες Σχεδιασμού του Ε.Π. «Ψηφιακή Σύγκλιση», <http://www.opengov.gr/ypoiar/?option=sygglisi>.
- [9] Ouksel A.M. and Sheth A. (1999) Semantic Interoperability in Global Information Systems, ACM SIGMOD Record, Vol 28(1) March 1999, pp 5-12.
- [10] Europeana portal, <http://www.europeana.eu>
- [11] DRIVER portal, <http://www.driver-repository.eu/>
- [12] Shreeves, S. L., J. Riley, and L. Milewicz (2006, August). Moving towards shareable metadata. *First Monday* 11(8).
- [13] Angelaki, G., R. Caffo, M. Hagedorn-Saupe, and S. Hazan (2010, April). Athena: A mechanism for harvesting europe's museum holdings into europeana. In *Museum and the Web 2010*, <http://www.archimuse.com/mw2010/papers/angelaki/angelaki.html>.
- [14] Athena project (2009, April). Deliverable 3.1: Report on existing standards applied by european museums, <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>.
- [15] Athena project (2009, July). Deliverable 3.2: Recommendations and best practice report, <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>.
- [16] (2008, November). «Ελληνικό Πλαίσιο Παροχής Υπηρεσιών Ηλεκτρονικής Διακυβέρνησης και Πρότυπα Διαλειτουργικότητας», Πλαίσιο Πιστοποίησης Δημόσιων Διαδικτυακών Τόπων, διαθέσιμο στο <http://www.e-gif.gov.gr/portal/pls/portal/docs/216024.PDF>
- [17] DRIVER Guidelines 2.0: Guidelines for content providers - Exposing textual resources with OAI-PMH, Νοέμβριος 2008. Διαθέσιμο στη διεύθυνση http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf.

- [18] DINI Certificate: Document and Publication Services 2007, Τεχνική αναφορά, έκδοση 2.0. Διαθέσιμο στη μόνιμη διεύθυνση <http://nbn-resolving.de/urn:nbn:de:kobv:11-10075687>.
- [19] The ISO/IEC 27000-series numbering ("ISO27k") has been reserved for a family of information security management standards <http://www.iso27001security.com/html/iso27000.html>
- [20] BS 7799 <http://www.bsigroup.com/>, British Library Digital Preservation Team – strategy διαθέσιμο στο <http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresteamstrat/index.html>
- [21] Office of Government Commerce (2002). ICT Infrastructure Management. The Stationery Office. ISBN 0113308655.
- [22] OAIS, Open archival information system, ISO 14721:2003
- [23] Trustworthy Repositories Audit and Certification Checklist (TRAC), http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf
- [24] Blue ribbon task force Sustainable economics for digital planet ensuing long term access to Digital Information http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- [25] The Handle System, <http://www.handle.net/>
- [26] «Ελληνικό Πλαίσιο Παροχής Υπηρεσιών Ηλεκτρονικής Διακυβέρνησης και Πρότυπα Διαλειτουργικότητας», Πλαίσιο Πιστοποίησης Δημόσιων Διαδικτυακών Τόπων, διαθέσιμο στο <http://www.e-gif.gov.gr/portal/pls/portal/docs/216024.PDF>

Κατάλογος Πινάκων και Εικόνων

Πίνακες

Πίνακας 1. Ελάχιστες απαιτήσεις για ψηφιοποίηση διαφορετικών αντικειμένων σύμφωνα με [1]	9
Πίνακας 2. Επιπρόσθετες απαιτούμενες προδιαγραφές ψηφιοποίησης.....	11
Πίνακας 3. Επίπεδα διαλειτουργικότητας αποθετηρίων περιεχομένου.....	14
Πίνακας 4. Προδιαγραφές για όλες τις κατηγορίες αποθετηρίων (Επίπεδο 1)	18
Πίνακας 7 Προδιαγραφές για απαίτηση μόνιμων προσδιοριστών ψηφιακών πόρων	19
Πίνακας 8. Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού.....	20
Πίνακας 9. Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού.....	21
Πίνακας 10. Προδιαγραφές παρουσίασης και διάθεσης για κείμενα	21
Πίνακας 11. Προδιαγραφές παρουσίασης και διάθεσης για εικόνες	22
Πίνακας 12. Διαλειτουργικότητα με μητρώα ανοικτού ψηφιακού περιεχομένου και σύστημα ασφαλούς διαφύλαξης ανοικτού ψηφιακού περιεχομένου	23
Πίνακας 13. Ενδεικτική υλοποίηση συστημάτων διάθεσης ψηφιακού περιεχόμενο με ΕΛ/ΛΑΚ και σχετικά πρότυπα.....	25
Πίνακας 14. Σύνοψη απαιτούμενων λειτουργικών χαρακτηριστικών έργων περιεχομένου	27

Εικόνες

Εικόνα 1 Απλουστευμένη μορφή ψηφιακού τεκμηρίου και σχετικών προτύπων	7
Εικόνα 2 Ψηφιακό περιεχόμενο, διαλειτουργικότητα και διατήρησή του.	12
Εικόνα 3 Ενδεικτική «στοίβα» ΕΛ/ΛΑΚ για διάθεση ψηφιακού περιεχομένου σύμφωνα.....	25



EKT

ΕΘΝΙΚΟ ΚΕΝΤΡΟ
ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL
DOCUMENTATION
CENTRE

www.epset.gr

Εθνικό Πληροφοριακό Σύστημα
Έρευνας και Τεχνολογίας

Κοινωνικά Δίκτυα - Περιεχόμενο Παραγόμενο από Χρήστες