



ΕΘΝΙΚΟ ΚΕΝΤΡΟ
ΤΕΚΜΗΡΙΩΣΗΣ

N A T I O N A L
DOCUMENTATION
C E N T R E

**Καλές πρακτικές και προδιαγραφές
διαλειτουργικότητας για ανοικτό τεκμηριωμένο
ψηφιακό περιεχόμενο**

Περιεχόμενα

Περιεχόμενα.....	2
Σκοπός.....	3
Αποθετήρια για την οργάνωση, διάθεση και διατήρηση τεκμηριωμένου ψηφιακού περιεχομένου	5
Καλές πρακτικές και προδιαγραφές διαλειτουργικότητας για την ψηφιοποίηση, οργάνωση και διάθεση του πολιτιστικού αποθέματος	7
Ψηφιοποίηση	7
Μεταδεδομένα και διαλειτουργικότητα	10
Μόνιμοι προσδιοριστές (διευθύνσεις) πόρων ψηφιακού πολιτιστικού αποθέματος.....	15
Διάθεση τεκμηριωμένου ψηφιακού περιεχομένου.....	16
Διαλειτουργικότητα με μητρώο ανοικτού τεκμηριωμένου ψηφιακού περιεχομένου και βασικά προσ απαιτούμενα ασφάλειας και ψηφιακής διατήρησης.....	18
Λογισμικό για την ανάπτυξη υποδομών διάθεσης ψηφιακού περιεχομένου.....	20
Σύνοψη επιθυμητών χαρακτηριστικών και προδιαγραφών.....	22
Βιβλιογραφία.....	24
Κατάλογος Πινάκων και Εικόνων.....	26
Πίνακες.....	26
Εικόνες.....	26

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 δημοσίων τεκμηρίωσης και ψηφιακού περιεχομένου, σύμφωνα και με τις απαιτήσεις για την Ηλεκτρονική Διακυβέρνηση.
- δεν λαμβάνει υπόψη τις νέες τεχνολογικές αποδοτικά προσεγγίσεις διάθεσης εφαρμογών και Cloud γενικά και Software as a Service Cloud υποδομών, για 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 τους σε παραγωγική λειτουργία ανά μεμονωμένη εγκατάσταση,
- τις μειωμένες δυνατότητες επαναχρησιμοποίησής του,
- 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, αξιοποίησής και ασφάλειας του.

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

Παρότι κατά την προηγούμενη προγραμματική περίοδο είχαν δοθεί στους αναδόχους έργων με τη μορφή μελετών 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.

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, πρέπει να διαμορφωθεί ένα πλαίσιο για την τήρηση των προδιαγραφών που πρέπει να πληρούν τα συστήματα διάθεσης ψηφιακού περιεχομένου το οποίο ταυτόχρονα θα εξασφαλίζει τα απαραίτητα σημεία ελέγχου 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 έχουν προδιαγραφεί οι κύριες λειτουργικές προδιαγραφές και οι καλές πρακτικές που θα πρέπει να υποστηρίζονται, κατά ελάχιστο από το σύστημα αποθετηρίου που χρησιμοποιείται για την οργάνωση, απόθεση και διάθεσης του ψηφιακού πολιτιστικού αποθέματος.

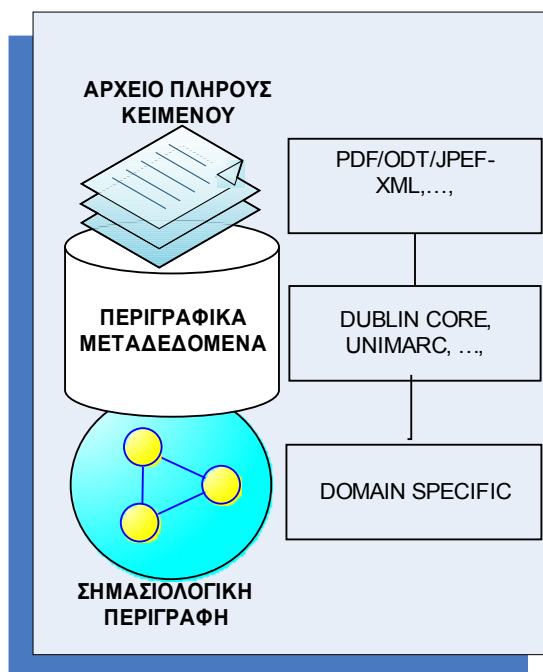
Για την επίτευξη βιωσιμότητας, την επίτευξη οικονομιών κλίμακας, διατίθεται προς του φορείς υποδομή Αποθετηρίου, η οποία συμμορφώνεται με αυτές τις διεθνείς καλές πρακτικές και προδιαγραφές, παρεχόμενη από το Εθνικό Κέντρο Τεκμηρίωσης (ΕΚΤ) με την μορφή Software as a Service Cloud. Εναλλακτικά συστήματα διάθεσης, εφόσον κρίνονται απαραίτητα, θα πρέπει να συμμορφώνονται σε αυτές τις διεθνείς καλές πρακτικές και προδιαγραφές.

Επιπλέον, για την δημιουργία ενιαίου ψηφιακού πολιτιστικού αποθέματος και για να διασφαλισθεί η διατήρηση και βιώσιμότητα των ψηφιακών πόρων, απαιτείται η απόθεση του ψηφιακού πολιτιστικού αποθέματος και των αντιστοίχων μεταδεδομένων, σε μητρώο ανοικτού τεκμηριωμένου ψηφιακού περιεχομένου παρεχόμενο από το Εθνικό Κέντρο Τεκμηρίωσης, με βάση τα διαδεδομένα διεθνή πρότυπα διαλειτουργικότητας στον αντίστοιχο χώρο.

Αποθετήρια για την οργάνωση, διάθεση και διατήρηση τεκμηριωμένου 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 διαφέρουν ποικίλουν ανάλογα με το είδος, μοιράζονται ωστόσο τις βασικές αρχές διαλειτουργικότητας και ανάγκες παρουσίασης και διάθεσης του περιεχομένου.

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

Καλές πρακτικές και προδιαγραφές διαλειτουργικότητας για την ψηφιοποίηση, οργάνωση και διάθεση του πολιτιστικού αποθέματος

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», και ενσωματώνουν υλικό από αντίστοιχες διαδεδομένες διεθνείς μελέτες και πηγές. Ιδιαίτερη σημασία έχουν οι ελάχιστες απαιτήσεις ποιότητας που, σύμφωνα με την προαναφερόμενη μελέτη, είναι:

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

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 εφόσον λαμβάνονται υπόψη **είναι επαρκείς**, αν και πρέπει να εξεταστεί επικαιροποίηση τους όπως π.χ με την προσθήκη της δυνατότητας για αποθήκευση των εικόνων και σε μορφή lossless JPEG 2000 κ.α..

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 το μητρώο ανοικτού τεκμηριωμένου περιεχομένου, η αναζήτηση και το indexing στο πλήρες κείμενο (fulltext 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 τεκμηριωμένου ψηφιακού περιεχομένου.

Για την επεξεργασία των ψηφιοποιημένων αρχείων για το σύνολο των έργων πολιτιστικού 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.

C. 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.

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

¹ Αφορά κείμενα από τυπογραφικά δοκίμια και όχι χειρόγραφα κείμενα.

A/A	Απαίτηση	Περιγραφή
1	Οπτική Αναγνώριση Χαρακτήρων (OCR) για υποστήριξη αναζήτησης και ευρετηρίασης πλήρους κειμένου του	<p>Παραγωγή αδιόρθωτου OCR για το σύνολο του προς ψηφιοποίηση έντυπου περιεχομένου στις παρακάτω μορφές:</p> <ul style="list-style-type: none"> ▪ Παράδοση τελικών PDF αρχείων κατά τρόπο ώστε πίσω από την ψηφιοποιημένη εικόνα να υπάρχει το αντικείμενο του σε «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. ▪ 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. ▪ Σε αναλυτικό αρχείο στις πρότυπες μορφές μορφή ABBYY XML ή 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.

Πίνακας 2. Επιπρόσθετες απαιτούμενες προδιαγραφές ψηφιοποίησης

Μεταδεδομένα και διαλειτουργικότητα

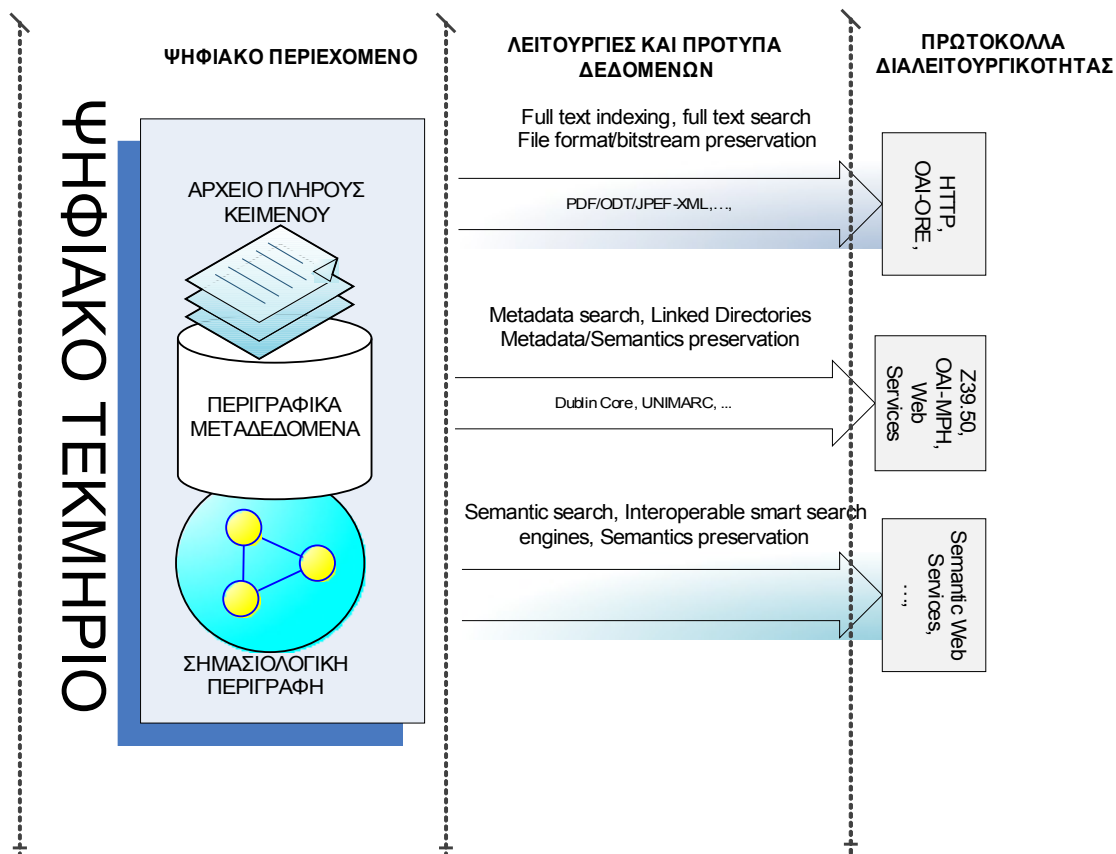
Ιδιαίτερη συνάφεια με τη διαλειτουργικότητα σε επίπεδο μεταδεδομένων έχουν οι έννοιες της διαλειτουργικότητας και των ανοικτών δεδομένων, οι οποίες ορίζονται στην πρόσφατη [Δημόσια διαβούλευση για τις γενικές αρχές και προτεραιότητες Σχεδιασμού του Ε.Π. «Ψηφιακή Σύγκλιση»](#) :

- **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 κατατεταγμένα, 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 :

- **Διαλειτουργικότητα σε επίπεδο αποθετηρίων και μητρώου ανοικτού τεκμηριωμένου περιεχομένου.** Επιτρέπει την online διάθεση των μεταδεδομένων ενός αποθετηρίου σε τρίτες εφαρμογές και συστήματα, μέσω υποσυστήματος που 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 διαμοιρασμού (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.

Η επίτευξη της ιδανικής διαλειτουργικότητας σε όλα τα επίπεδα αποτελεί εξαιρετικά δύσκολο εγχείρημα, ιδιαίτερα όταν αφορά μεγάλο πλήθος αναγκαστικά ετερογενούς πολιτιστικού αποθέματος. 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).

▪ **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.

▪ **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 ένα ικανοποιητικό επίπεδο σημασιολογικής συμβατότητας.

Επίπεδο	Εύρος εφαρμογής	Περιγραφή	Υποστηριζόμενες υπηρεσίες
1	Όλα τα επιμέρους αποθετήρια	Διάθεση μέσω web services για harvesting των μεταδεδομένων (τρία πεδία του unqualified Dublin Core: title, type, identifier). Υποστήριξη μοναδικών προσδιοριστών ανά εγγραφή μεταδεδομένων. Ευρετηρίαση τουλάχιστον των μεταδεδομένων από τις κύριες διαδικτυακές μηχανές αναζήτησης γενικού σκοπού. Κωδικοποίηση τιμών πεδίων σύμφωνα με διεθνή πρότυπα. Διάθεση πληροφορίας για την κατάσταση του περιεχομένου όσον αφορά θέματα πνευματικής ιδιοκτησίας.	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	Όλα τα αποθετήρια μιας κατηγορίας	Διάθεση μέσω web services για harvesting των μεταδεδομένων και μετα-αναζήτηση. Υποστήριξη ευρύτερου υποσυνόλου του unqualified Dublin Core καθώς και του Europeana Data Model. Προσδιορισμός συγκεκριμένων sets για επιλεκτικό harvesting. Διάθεση μεταδεδομένων ως linked data. Κωδικοποίηση τιμών πεδίων σύμφωνα με διεθνή πρότυπα.	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	Τα μεγαλύτερα και πιο σημαντικά αποθετήρια ανά κατηγορία	Υποστήριξη αναλυτικού σχήματος μεταδεδομένων. Κοινά εργαλεία για σημασιολογική συμβατότητα (π.χ. θεματικές επικεφαλίδες, γεωγραφική πληροφορία).	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.

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 και DRIVER . 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 ³). Παράλληλα, προβλέπεται και η κωδικοποίηση των λεπτομερών (ανά κατηγορία υλικού) προτύπων μεταδεδομένων σύμφωνα με το Europeana Data Model

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 .

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

Προδιαγραφές για όλες τις κατηγορίες αποθετηρίων (Επίπεδο 1)

A/A	Απαιτήση	Προδιαγραφή
	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 αναζήτησης	Πρέπει να υποστηρίζεται η ευρετηρίαση τουλάχιστον των μεταδεδομένων από συνήθεις διαδικτυακές μηχανές αναζήτησης (Google, Bing, Yahoo).
4.	Διαλειτουργικότητα σε επίπεδο συστημάτων ψηφιακών βιβλιοθηκών και καταλόγων	Θα πρέπει να υποστηρίζεται ανάκτηση των μεταδεδομένων βάσει του πρωτοκόλλου OAI-PMH, έκδοση 2.0. Η συγκεκριμένη λειτουργία θα πρέπει να διατίθεται δημόσια, χωρίς περιορισμούς πρόσβασης.

A/A	Απαίτηση	Προδιαγραφή
5.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής μεταδεδομένων και δεδομένων	<ul style="list-style-type: none"> Στα Unqualified Dublin Core μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να περιέχεται ένας τουλάχιστον μόνιμος προσδιοριστής (persistent identifiers) που να ταυτοποιεί την εγγραφή. 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. 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). Η κωδικοποίηση του πεδίου dc.date (αν υπάρχει) στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να ακολουθεί το πρότυπο ISO 8601.

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

Προδιαγραφές για αποθετήρια κειμένων (Επίπεδο 2)

A/A	Απαίτηση	Προδιαγραφή
	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής μεταδεδομένων	Πέντε πεδία του Unqualified Dublin Core θα πρέπει να είναι υποχρεωτικά για κάθε εγγραφή: title, creator, date, type, identifier.
2.	Διαλειτουργικότητα σε επίπεδο συστημάτων ψηφιακών βιβλιοθηκών και καταλόγων	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.
3.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής μεταδεδομένων Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Στα Unqualified Dublin Core μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να υπάρχει μόνο ένα πεδίο dc.date το οποίο θα πρέπει να αφορά την ημερομηνία έκδοσης του τεκμηρίου.
4.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής μεταδεδομένων	Η κωδικοποίηση του πεδίου dc.creator στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να ακολουθεί το πρότυπο βιβλιογραφικών αναφορών της American Psychological Association (APA) σε ό,τι αφορά τη γραφή των ονομάτων συγγραφέων.
5.	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής μεταδεδομένων Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Η κωδικοποίηση του πεδίου dc.type στα μεταδεδομένα εγγραφής που διατίθενται μέσω OAI-PMH θα πρέπει να είναι σύμφωνη με τις αντίστοιχες οδηγίες όπως έχουν αποτυπωθεί στα DRIVER Guidelines (, σελ. 68-71).

A/A	Απαίτηση	Προδιαγραφή
6.	Διαλειτουργικότητα σε επίπεδο συστημάτων αναζήτησης και καταλόγων	Θα πρέπει να ορίζονται με τη μορφή OAI-PMH Sets και γίνονται διαθέσιμα για ανάκτηση μέσω OAI-PMH σε μορφή Unqualified Dublin Core υποσύνολα των εγγραφών του αποθετηρίου, κατ' ελάχιστον τα ακόλουθα: <ul style="list-style-type: none"> i. Το σύνολο των εγγραφών που περιέχουν πλήρες κείμενο. ii. Από ένα σύνολο εγγραφών για κάθε διαφορετική τιμή του πεδίου dc.type, από αυτές που ακολουθούν την κωδικοποίηση για τύπους δημοσίευσης του (σελ. 69).

Πίνακας 5. Προδιαγραφές για αποθετήρια κειμένων (Επίπεδο 2)

Προδιαγραφές για αποθετήρια κειμένων (Επίπεδο 3)

A/A	Απαίτηση	Προδιαγραφή
	Διαλειτουργικότητα σε επίπεδο σύνταξης και δομής.	Τα μεταδεδομένα του αποθετηρίου είναι διαθέσιμα για ανάκτηση (harvesting) μέσω του πρωτοκόλλου OAI-PMH, έκδοση 2.0 σε μορφή κάποιου πιο αναλυτικού σχήματος αναπαράστασης μεταδεδομένων όπως Qualified Dublin Core, MODS, MARCXML, CERIF.
2.	Διαλειτουργικότητα σε επίπεδο συστημάτων αναζήτησης και καταλόγων. Διαλειτουργικότητα σε σημασιολογικό επίπεδο.	Τα μεταδεδομένα του αποθετηρίου είναι διαθέσιμα ως Linked Data.

Πίνακας 6. Προδιαγραφές για αποθετήρια κειμένων (Επίπεδο 3)

Το σύστημα Αποθετηρίου το οποίο θα παρέχεται με το μοντέλο SaaS από το Εθνικό Κέντρο Τεκμηρίωσης θα υποστηρίζει τις παραπάνω μορφές διαλειτουργικότητας. Για την πλήρη αξιοποίηση της επένδυσης σε περίπτωση που δεν γίνεται χρήση των παραπάνω διαθέσιμων συστημάτων θα πρέπει να απαιτούνται αντίστοιχες προδιαγραφές από τα προσφερόμενα συστήματα οργάνωσης και διάθεσης του ψηφιακού περιεχομένου.

Μόνιμοι προσδιοριστές (διευθύνσεις) πόρων ψηφιακού πολιτιστικού αποθέματος

Απαραίτητο για την ύπαρξη διαρκούς διαλειτουργικότητας είναι η σταθερή διευθυνσιοδότηση του κάθε ψηφιακού πόρου, ανεξάρτητα από το συγκεκριμένο σύστημα λογισμικού που χρησιμοποιείται για την διάθεση του ώστε να εξασφαλίζεται η πρόσβαση σε αυτόν ανεξάρτητα χρονικού διαστήματος και συστήματος λογισμικού που τον παρέχει. Απαιτείται παροχή των τεκμηρίων με αποθετήριο που να υποστηρίζει μηχανισμό persistent identifier handle service για την παροχή μονίμων τοποθεσιών ψηφιακού περιεχομένου. 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 διευθύνσεως ανά ψηφιακό πόρο.

1	Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier)	<ul style="list-style-type: none"> ▪ Έκδοση persistent identifiers από κατάλληλη αρχή ▪ Διάθεση ψηφιακού περιεχομένου στις
---	-----------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------

	handle service)	<p>διευθύνσεις που ορίζουν τα persistent identifiers</p> <ul style="list-style-type: none"> ▪ Τα persistent identifiers θα πρέπει να είναι μόνιμα ανεξάρτητα από σύστημα λογισμικού ή/και αποθετηρίου και να μην επηρεάζονται από σχετικές μεταβάσεις
--	------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Πίνακας 7 Προδιαγραφές για απαίτηση μόνιμων προσδιοριστών ψηφιακών πόρων

Διάθεση τεκμηριωμένου 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 .

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 πλήρεις 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.

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

A/A	Απαίτηση	Προδιαγραφή
	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	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

	εξωτερικές πηγές.	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-σύνδεσμοι στις σελίδες παρουσίασης των αντίστοιχων τεκμηρίων.

Πίνακας 8. 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
	Πολυκαναλική διάθεση	Διαθεσιμότητα του περιεχομένου – με κατάλληλες προσαρμογές – για προβολή σε συσκευές εναλλακτικές των συνήθων desktop / laptop υπολογιστών όπως e-book readers, net-books, tablets, smart phones.

Πίνακας 9. Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού (Επίπεδο 2)

Προδιαγραφές παρουσίασης και διάθεσης για κείμενα

A/A	Απαίτηση	Προδιαγραφή
	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.
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 10. Προδιαγραφές παρουσίασης και διάθεσης για κείμενα

Προδιαγραφές παρουσίασης και διάθεσης για εικόνες

A/A	Απαίτηση	Προδιαγραφή
	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.

A/A	Απαίτηση	Προδιαγραφή
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 συνίσταται, 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 11. Προδιαγραφές παρουσίασης και διάθεσης για εικόνες

Το Αποθετήριο παρεχόμενο από το Εθνικό Κέντρο Τεκμηρίωσης ως SaaS θα υποστηρίζει τα παραπάνω. Για την πλήρη αξιοποίηση της επένδυσης σε περίπτωση που δεν γίνεται χρήση της υποδομής SaaS θα πρέπει να απαιτούνται αντίστοιχες προδιαγραφές από τα συστήματα οργάνωσης και διάθεσης του ψηφιακού περιεχομένου.

Διαλειτουργικότητα με μητρώο ανοικτού τεκμηριωμένου ψηφιακού περιεχομένου και βασικά προαπαιτούμενα ασφάλειας και ψηφιακής διατήρησης

Το μητρώο ανοικτού περιεχομένου ψηφιακού περιεχομένου λειτουργεί για την ευρετηρίαση του τεκμηριωμένου ψηφιακού περιεχομένου, με σκοπό την δημιουργία ενιαία αναζητήσιμου ψηφιακού αποθέματος, καθώς και για την εξασφάλιση της εφαρμογής βασικών μηχανισμών ασφάλειας και ψηφιακής διατήρησης του ανοικτού ψηφιακού τεκμηριωμένου περιεχομένου που έχει αναπτυχθεί.

Το ψηφιακό τεκμηριωμένο απόθεμα και τα αντίστοιχα μεταδεδομένα, θα συλλέγονται, για ευρετηρίαση στο μητρώο ανοικτού τεκμηριωμένου ψηφιακού περιεχομένου του Εθνικού Κέντρου Τεκμηρίωσης, προκειμένου να επιτευχθεί η δημιουργία ενιαίου ψηφιακού πολιτιστικού αποθέματος και να διασφαλισθεί η διατήρηση και βιώσιμότητα των ψηφιακών πόρων.

Το μητρώο θα συλλέγει – με περιοδική αυτόματη ανανέωση σε τακτά χρονικά διαστήματα – τα μεταδεδομένα τεκμηρίωσης των επιμέρους αποθετηρίων. Συμπεριλαμβάνεται, για λόγους ασφάλειας και απομακρυσμένης αντιγραφής των δεδομένων με την ύπαρξη ενός αντιγράφου του ψηφιακού περιεχομένου στο Εθνικό Κέντρο Τεκμηρίωσης, και την εφαρμογή βασικών πολιτικών διατήρησης και το ίδιο το ψηφιακό υλικό.

Επιπλέον το σύστημα του μητρώου θα προσφέρει λειτουργίες ασφάλειας δεδομένων και βασικής ψηφιακής διατήρησης, μέσω της συλλογής των πλήρων ψηφιακών πόρων, στο μητρώο ανοικτού τεκμηριωμένου περιεχομένου του ΕΚΤ. Διασφαλίζεται ένα τουλάχιστον **απομακρυσμένο αντίγραφο** των δεδομένων, μεταδεδομένων και του πρωτότυπου ψηφιακού περιεχομένου λειτουργώντας ως το **disaster recovery site** στο τομέα του ψηφιακού περιεχομένου, αξιοποιώντας ήδη διαθέσιμα χαρακτηριστικά και διεπαφές των αποθετηρίων (οργανωμένα μεταδεδομένα, διεπαφές, μοναδικοί αναγνωριστές) για τη διασφάλιση βασικών λειτουργιών remote data replication και disaster recovery με μειωμένο κόστος.

A/A	Απαίτηση	Περιγραφή
1.	Εναπόθεση μεταδεδομένων αποθετηρίων στο μητρώο ανοικτού τεκμηριωμένου περιεχομένου	Παροχή μηχανισμών ανταλλαγής δεδομένων και μεταδεδομένων με τα πρότυπα: <ul style="list-style-type: none"> ▪ ΟΑΙ-ΡΜΗ, ▪ Περιοδική συλλογή μεταδεδομένων ▪ Ευρετηρίαση μεταδεδομένων
2.	Συλλογή ψηφιακών πόρων αποθετηρίων στο μητρώο ανοικτού τεκμηριωμένου περιεχομένου	<ul style="list-style-type: none"> ▪ ΟΑΙ-ΟΡΕ ή custom batch μηχανισμοί για μεταφορά ▪ Handle Service και Persistent identifiers ▪ Εξασφάλιση remote replication πόρων και βασικών χαρακτηριστικών διατήρησης

Πίνακας 12. Προδιαγραφές έργων ψηφιακών βιβλιοθηκών/αποθετηρίων για ψηφιακή διατήρηση και ασφάλεια περιεχομένου

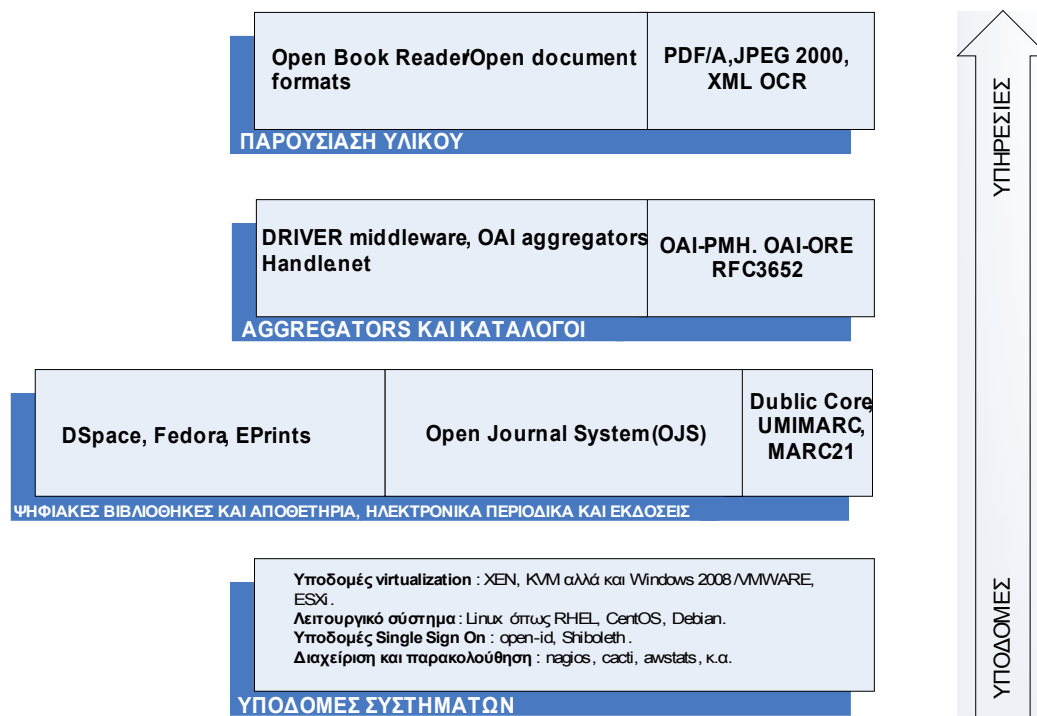
Λογισμικό για την ανάπτυξη υποδομών διάθεσης ψηφιακού περιεχομένου

Στον Πίνακα 13 αναφέρονται ενδεικτικά συστήματα ΕΛ/ΛΑΚ τα οποία μπορούν να αξιοποιηθούν για την πλήρη υλοποίηση έργων ψηφιακών αποθετηρίων. Αν και ορισμένα από αυτά αποτελούν ιδιαίτερα διαδεδομένα συστήματα, η αναφορά γίνεται ώστε 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» () 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
	Παρουσίαση ψηφιακών τεκμηρίων	Λογισμικά ΕΛ/ΛΑΚ <ul style="list-style-type: none"> Internet Archive Book Reader Flexpaper 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 Greenstone OJS ABEKT (νέα έκδοση) 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	Υπηρεσίες υποδομής: Λειτουργικά Συστήματα, Middleware/B.Δ. Πλατφόρμες virtualization Συστήματα Single Sign On (SSON) Συστήματα διαχείρισης και παρακολούθησης	<ul style="list-style-type: none"> Linux (RHEL, CentOS, ...) Tomcat, JBoss, PostgreSQL, MySQL XEN, KVM ακόμα και Windows 2008/VMWARE, ESXi μιας και δρουν πολλαπλασιαστικά Shibboleth, openid nagios, cacti, awstats, κ.α.

Πίνακας 13. Υλοποίηση συστημάτων διάθεσης ψηφιακού περιεχόμενο με ΕΛ/ΛΑΚ και σχετικά πρότυπα



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

Σύνοψη επιθυμητών 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 και 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			✓		✓
Μόνιμες διευθύνσεις			✓	✓	✓

	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)
ψηφιακών τεκμηρίων (persistent identifier handle service)					
Μόνιμες διευθύνσεις ψηφιακών τεκμηρίων (persistent identifier handle service)			✓	✓	✓

Πίνακας 14. 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	7
Πίνακας 2. Επιπρόσθετες απαιτούμενες προδιαγραφές ψηφιοποίησης.....	9
Πίνακας 3. Επίπεδα διαλειτουργικότητας αποθετηρίων περιεχομένου.....	12
Πίνακας 4. Προδιαγραφές για όλες τις κατηγορίες αποθετηρίων (Επίπεδο 1).....	14
Πίνακας 5. Προδιαγραφές για αποθετήρια κειμένων (Επίπεδο 2).....	15
Πίνακας 6. Προδιαγραφές για αποθετήρια κειμένων (Επίπεδο 3).....	15
Πίνακας 7 Προδιαγραφές για απαίτηση μόνιμων προσδιοριστών ψηφιακών πόρων	16
Πίνακας 8. Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού	17
Πίνακας 9. Προδιαγραφές παρουσίασης και διάθεσης για όλους τους τύπους υλικού	17
Πίνακας 10. Προδιαγραφές παρουσίασης και διάθεσης για κείμενα.....	17
Πίνακας 11. Προδιαγραφές παρουσίασης και διάθεσης για εικόνες.....	18
Πίνακας 12. Προδιαγραφές έργων ψηφιακών βιβλιοθηκών/αποθετηρίων για ψηφιακή διατήρηση και ασφάλεια περιεχομένου.....	19
Πίνακας 13. Υλοποίηση συστημάτων διάθεσης ψηφιακού περιεχόμενο με ΕΛ/ΛΑΚ και σχετικά πρότυπα.....	20
Πίνακας 14. Σύνοψη απαιτούμενων λειτουργικών χαρακτηριστικών έργων περιεχομένου .	23

Εικόνες

Εικόνα 1 Απλουστευμένη μορφή ψηφιακού τεκμηρίου και σχετικών προτύπων.....	5
Εικόνα 2 Ψηφιακό περιεχόμενο, διαλειτουργικότητα και διατήρησή του.....	11
Εικόνα 3 Ενδεικτική «στοίβα» ΕΛ/ΛΑΚ για διάθεση ψηφιακού περιεχομένου σύμφωνα ...	21