



ΕΘΝΙΚΟ ΚΕΝΤΡΟ
ΤΕΚΜΗΡΙΩΣΗΣ

NATIONAL
DOCUMENTATION
CENTRE



Προδιαγραφές Ψηφιοποίησης και Ψηφιακών Αρχείων



Validator

πιστοποίηση προδιαγραφών
διαλειτουργικότητας & περιεχομένου



ΜΑΪΟΣ 2016

ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ

Εθνικό Κέντρο Τεκμηρίωσης | ΕΙΕ

1^η ΕΚΔΟΣΗ / Μάιος 2016

Μονάδα Ανάπτυξης Εφαρμογών ΕΚΤ

Κείμενα: Ιωάννα-Ουρανία Σταθοπούλου, PhD

Σχόλια & Προσθήκες: Παρασκευάς Καμάτσος (Ενότητα 1 – Εισαγωγή), Ιωάννης Βουλγαράκης (Ενότητα 2.1), Ιωάννης Δεσύπρης (Ενότητες 2.4 & 2.5),

Σχεδιασμός Έκδοσης: Δήμητρα Πελεκάνου

Επιμέλεια & Οπτικοποίηση Ύλης: Ιωάννα-Ουρανία Σταθοπούλου, Δήμητρα Πελεκάνου

Copyright © 2016 Εθνικό Κέντρο Τεκμηρίωσης | ΕΙΕ

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

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



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

Προκειμένου να δείτε αντίγραφο της άδειας επισκεφθείτε:
<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.el>



πιστοποίηση προδιαγραφών
διαλειτουργικότητας & περιεχομένου

ΠΕΡΙΕΧΟΜΕΝΑ

Σχετικά με τον οδηγό.....	1
1 Βασικές έννοιες - διαδικασίες	3
1.1 Παραγωγή ψηφιακού αρχείου	5
1.2 Σύστημα Διαχείρισης και Διάθεσης Περιεχομένου	5
1.3 Τεκμηρίωση και μεταδεδομένα.....	6
1.4 Ανοικτά Διασυνδεδεμένα δεδομένα	7
1.5 Συστήματα συσσώρευσης περιεχομένου	7
1.6 Επίπεδα διαλειτουργικότητας	8
2 Προδιαγραφές για ψηφιακά αρχεία.....	10
2.1 Τύποι περιεχομένου EKT και ομάδες μορφότυπων αρχείων	10
2.2 Προδιαγραφές για αρχεία εικόνας	11
2.3 Προδιαγραφές για αρχεία κειμένου	18
2.4 Προδιαγραφές για αρχεία βίντεο	23
2.5 Προδιαγραφές για αρχεία ήχου	29
2.6 Προδιαγραφές για σύνολα δεδομένων	33
3 Βιβλιογραφία	35
4 Κατάλογος Πινάκων και Εικόνων	37
4.1 Κατάλογος Πινάκων	37
4.2 Κατάλογος Εικόνων	37

Σχετικά με τον οδηγό

Το ΕΚΤ εκπληρώνοντας τον θεσμικό του ρόλο ως επιστημονική υποδομή εθνικής χρήσης που συλλέγει, οργανώνει και διαθέτει επιστημονικό και πολιτιστικό περιεχόμενο, εφαρμόζει μία σταθερή πολιτική συνεργασιών με επιστημονικούς και πολιτιστικούς φορείς με σκοπό την παραγωγή και διάθεση έγκριτου ψηφιακού περιεχομένου προσβάσιμο από όλους. Στο πλαίσιο αυτό αναπτύσσει υπηρεσίες που υποστηρίζουν φορείς όπως μουσεία, αρχεία, βιβλιοθήκες και ερευνητικά κέντρα να οργανώσουν και να διαθέσουν το περιεχόμενο τους μέσα από κατάλληλες υποδομές και την υλοποίηση προτύπων και προδιαγραφών ώστε να διευκολύνεται η κυκλοφορία και επανάχρηση του.

Το ΕΚΤ είχε συντάξει τη μελέτη «Προδιαγραφές και χαρακτηριστικά διαλειτουργικότητας για ανοικτό ψηφιακό περιεχόμενο» [24], το οποίο περιελάμβανε τόσο τις προδιαγραφές διαλειτουργικότητας για τη διάθεση και τα μεταδεδομένα όσο και κάποιες συνοπτικές οδηγίες σχετικά με την ψηφιοποίηση. Η αρχική έκδοση της μελέτης και η 1^η και 2^η επικαιροποίηση της έλαβαν χώρα στο πλαίσιο του Έργου «Εθνικό Πληροφοριακό Σύστημα Έρευνας και Τεχνολογίας/Κοινωνικά Δίκτυα – Περιεχόμενο Παραγόμενο από Χρήστες», ενώ η 3^η επικαιροποίηση (2^η έκδοση) υλοποιήθηκε στο πλαίσιο του έργου «Πλατφόρμα Παροχής Υπηρεσιών Κατάθεσης, Διαχείρισης και Διάθεσης Ανοικτών Δεδομένων και Ψηφιακού Περιεχομένου». Μέρος των προδιαγραφών της 2^{ης} έκδοσης, ενσωματώθηκαν στο πλαίσιο των Προσκλήσεων 31 και 31.2 του Επιχειρησιακού Προγράμματος «Ψηφιακή Σύγκλιση», όπου 46 Πολιτιστικοί Φορείς που εντάχθηκαν στο έργο, ψηφιοποίησαν και διέθεσαν το περιεχόμενο τους, σύμφωνα με τις προδιαγραφές που είχε θέσει το ΕΚΤ.

Ο παρών οδηγός «Προδιαγραφές ψηφιοποίησης και ψηφιακών αρχείων», περιγράφει τις προδιαγραφές και τις διαδικασίες που συστήνεται να ακολουθηθούν κατά τη διάρκεια παραγωγής ψηφιακών πόρων, ώστε να εξασφαλίζεται όσο το δυνατόν η μακροχρόνια διαφύλαξη τους και η διάθεση τους στους χρήστες. Περιλαμβάνει τους πιο δημοφιλείς τύπους ψηφιακού περιεχομένου, όπως κείμενο, εικόνα, ήχος, βίντεο κλπ., και αποτυπώνει τις προδιαγραφές τόσο για υλικό που έχει προκύψει από ψηφιοποίηση, αλλά και για εξ'αρχής ψηφιακό υλικό (digital born). Οι προδιαγραφές αυτές προέκυψαν από πρότυπα και καλές διεθνείς πρακτικές, καθώς και την εκτεταμένη εμπειρία του ΕΚΤ σε θέματα διαλειτουργικότητας και ψηφιακών αρχείων.

Οι Προδιαγραφές Ψηφιοποίησης και Ψηφιακών Αρχείων, ολοκληρώνονται με ένα ακόμη οδηγό, στον οποίο περιγράφονται οι Προδιαγραφές Διαλειτουργικότητας Ανοικτού Περιεχομένου με άλλα συστήματα και οι Προδιαγραφές Διάθεσης προς τον Χρήστη.

Οι προδιαγραφές, υποστηρίζουν την υπηρεσία Πιστοποίησης Προδιαγραφών Περιεχομένου του ΕΚΤ, μια υπηρεσία SaaS. Οι υπηρεσίες SaaS είναι ολοκληρωμένες ψηφιακές υπηρεσίες με στόχο την ενίσχυση της ψηφιακής παρουσίας των φορέων. Αξιοποιούν σύγχρονα επιχειρησιακά μοντέλα, ολοκληρωμένες λύσεις και υποδομές και επιτρέπουν την τεκμηριωμένη απόθεση, ασφαλή διαφύλαξη, οργάνωση και διάθεση ψηφιακού περιεχομένου. Βασίζονται σε τεχνολογίες αιχμής, όπως το μοντέλο SaaS (Software as a Service=Λογισμικό ως Υπηρεσία) και το Cloud Computing (Υπολογιστικό Νέφος), δηλαδή παρέχονται από το ΕΚΤ διαδικτυακά, χωρίς να χρειάζεται τοπική εγκατάσταση.

Η υπηρεσία Πιστοποίησης Προδιαγραφών Περιεχομένου προσφέρεται σε φορείς που επιθυμούν να ελέγξουν και να πιστοποιήσουν ότι το ψηφιακό τους περιεχόμενο πληροί ορισμένες προϋποθέσεις για να μπορεί να διατεθεί ανοικτά και ευρέως. Πρόκειται για μία υπηρεσία, η οποία παρέχεται μέσα από διαδικτυακή εφαρμογή Validator [<http://validator.ekt.gr>], την οποία διαχειρίζονται εξουσιοδοτημένοι χρήστες. Η υπηρεσία πραγματοποιεί μαζικούς αυτοματοποιημένους ελέγχους και παράγει αναλυτικές αναφορές υψηλού επιπέδου, οι οποίες πιστοποιούν την πληρότητα και την ποιότητα των μεταδεδομένων και των ψηφιακών αρχείων και διασφαλίζουν τις παρεχόμενες υπηρεσίες των συστημάτων διαχείρισης περιεχομένου των φορέων. Ο χρήστης έχει δυνατότητα να

παρακολουθεί όλη τη διαδικασία μέσα από τη διαδικτυακή πλατφόρμα. Το σύστημα είναι εύκολα επεκτάσιμο για νέα σχήματα μεταδεδομένων, άλλους τύπους ψηφιακών αρχείων και επιπρόσθετες προδιαγραφές διαλειτουργικότητας. Μέσω της πλατφόρμας έχουν ήδη πραγματοποιηθεί περισσότεροι από 200 έλεγχοι σε φορείς της Πρόσκλησης 31 και 31.2 ενώ νέοι φορείς προστίθενται συνεχώς στη διαδικασία ελέγχου.

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 επανάχρησης και διαφύλαξης του.

Σε ποιους απευθύνεται

- ➔ **Σε φορείς περιεχομένου:** Μουσεία, Φορείς Πολιτισμού, Δημόσιες και δημοτικές Βιβλιοθήκες, Πανεπιστήμια / ΑΤΕΙ, Ερευνητικά Κέντρα, Εκπαιδευτικούς Φορείς, Δημόσιους Φορείς για να ορίσουν τις προδιαγραφές των συστημάτων περιεχομένου τους στους αναδόχους που τα υλοποιούν.
- ➔ **Στο προσωπικό ψηφιοποίησης:** το οποίο ψηφιοποιεί και διαχειρίζεται πολλές φορές ευαίσθητα αντικείμενα ιστορικής και πολιτιστικής αξίας όπως βιβλία, πίνακες, παρτιτούρες, έργα γλυπτικής, κινηματογραφικό υλικό, χάρτες, χειρόγραφα, παλαίτυπα κ.α.
- ➔ **Στο προσωπικό τεκμηρίωσης:** το οποίο εμπλουτίζει τα ψηφιακά αρχεία με μεταδεδομένα εξυπηρετώντας τον εντοπισμό, τη διαχείριση των αντικειμένων και την παροχή πρόσβασης σε αυτά.
- ➔ **Στους προγραμματιστές:** οι οποίοι υλοποιούν τα συστήματα διαχείρισης και διάθεσης περιεχομένου, την πλατφόρμα πάνω στην οποία εργάζεται το προσωπικό του φορέα περιεχομένου και παράλληλα, διαθέτει αυτό τον όγκο πληροφορίας στον χρήστη και σε άλλα συστήματα συλλογής και διάθεσης περιεχομένου.

1 Βασικές έννοιες - διαδικασίες

Οι σύγχρονες εξελίξεις στις τεχνολογίες του παγκόσμιου ιστού ανοίγουν νέες δυνατότητες γύρω από τη χρήση της ηλεκτρονικής πληροφορίας. Πολιτιστικά, ιστορικά και επιστημονικά τεκμήρια των επιτευγμάτων του ανθρώπινου πνεύματος, όπως βιβλία, χειρόγραφα, παρτιτούρες, χάρτες, μουσειακά εκθέματα, πίνακες ζωγραφικής, έργα γλυπτικής, ηχητικό και κινηματογραφικό υλικό, **ψηφιοποιούνται** και καθίστανται διαθέσιμα σε ψηφιακή μορφή στον παγκόσμιο ιστό. Κύρια αποστολή των βιβλιοθηκών, των αρχείων και των μουσείων είναι η συλλογή, συντήρηση, διαχείριση και η **προβολή αυτής της πληροφορίας στο κοινό**.

Επιπλέον με την ψηφιοποίηση των πολιτιστικών και ιστορικών **τεκμηρίων εξυπηρετείται ο στόχος της μακροχρόνιας διατήρησης του περιεχομένου σε ψηφιακή μορφή αλλά και της διάθεσης του**, κάτι που είναι ζητούμενο και συζητείται σε διεθνές επίπεδο .

Παράλληλα όμως, οι πληροφορίες συλλέγονται από δικτυακές πύλες συσσώρευσης περιεχομένου, παρέχοντας τη δυνατότητα **ενιαίας αναζήτησης και οπτικοποίησης του**. Επομένως, σημαντικός παράγοντας σε αυτή τη διαδικασία είναι **η διαλειτουργικότητα του περιεχομένου**: τα δεδομένα αυτά θα πρέπει να διατίθενται με τέτοιο τρόπο ώστε να γίνονται κατανοητά τόσο από τους ανθρώπους όσο και από τις υπολογιστικές μηχανές.

Καθοριστική λειτουργία στο κύκλο ζωής του ψηφιακού περιεχομένου αποτελεί **η διαδικασία της τεκμηρίωσης**, της καταγραφής δηλαδή των χαρακτηριστικών και των ιδιοτήτων των αντικειμένων, με συστηματικό τρόπο, σχετικά με τη μορφή, την προέλευση και την ερμηνεία τους. Η διαδικασία της τεκμηρίωσης ακολουθεί κανόνες και διεθνή πρότυπα με στόχο την παραγωγή των μεταδεδομένων, ενός συνόλου πληροφορίας που είναι αναγνωρίσιμη και αξιοποιήσιμη από ανθρώπους και μηχανές. Η δημιουργία των μεταδεδομένων εντάσσεται στις κύριες εργασίες διαχείρισης και ανάδειξης του υλικού των φορέων και αποτελούν βασικό εργαλείο για την ουσιαστική αξιοποίηση του. Η εμπλοκή εξειδικευμένου προσωπικού στη διαδικασία της επεξεργασίας και τεκμηρίωσης του υλικού βελτιστοποιεί τις δυνατότητες επανάχρησής τους από συστήματα και υποδομές και κατ' επέκταση του περιεχομένου που περιγράφουν.

Ταυτόχρονα, η παραγωγή των μεταδεδομένων συμβάλλει σημαντικά στην ανάπτυξη έγκριτης επιστημονικής γνώσης για τα τεκμήρια, την προέλευση και την ερμηνεία τους και δημιουργεί έναν όγκο πληροφορίας που εμπλουτίζει το επιστημονικό και πολιτιστικό απόθεμα που διατίθεται στο κοινό των ειδικών και μη και ενεργοποιεί διαδικασίες έρευνας, εκπαίδευσης, πληροφόρησης και ψυχαγωγίας προς όφελος της κοινωνίας και των πολιτών.

Κεντρικός πυρήνας αυτής της διαδικασίας, είναι το **Σύστημα Διαχείρισης και Διάθεσης Περιεχομένου**, η πλατφόρμα στην οποία φιλοξενούνται τα ψηφιακά αρχεία και το προσωπικό του φορέα τεκμηριώνει και διαχειρίζεται όλο τον όγκο της πληροφορίας. Το σύστημα αυτό, παρέχει λειτουργίες προστιθέμενης αξίας για τη διάθεση του περιεχομένου στους χρήστες και **διαλειτουργεί με άλλα συστήματα, τα οποία συσσωρεύουν τα μεταδεδομένα και τα ψηφιακά αρχεία, παρέχοντας υπηρεσίες ενιαίας αναζήτησης και οπτικοποίησης**.

Η ολοκληρωμένη διαδικασία από την παραγωγή του ψηφιακού αρχείου, στην τεκμηρίωση και τη διάθεση προς τον χρήστη παρουσιάζεται στην Εικόνα 1.



1.1 Παραγωγή ψηφιακού αρχείου

Το πρώτο στάδιο της διαδικασίας είναι η παραγωγή του ψηφιακού αρχείου. Το αρχείο αυτό μπορεί να προκύψει από τα πρωτότυπα αντικείμενα (μη ψηφιακά) που μετατρέπονται σε ψηφιακά αντίγραφα (digitization), καθώς και από εξ' αρχής ψηφιακά αρχεία (digital born objects) ή ψηφιακά αντίγραφα (backup copy) αυτών [8]. Ωστόσο, με την τεχνολογία να εξελίσσεται με ταχύτατους ρυθμούς, το λογισμικό και οι μορφότυποι αλλάζουν πολύ γρήγορα, και ένα ψηφιακό αρχείο μπορεί να καταστεί ξεπερασμένο σε σχετικά σύντομο χρονικό διάστημα. Για το λόγο αυτό, η ψηφιοποίηση δεν ταυτίζεται αυτόματα με τη διαφύλαξη (preservation), καθώς τα ψηφιακά αρχεία που προκύπτουν από την ψηφιοποίηση χρειάζονται και αυτά συντήρηση. Σκοπός της ψηφιακής διαφύλαξης είναι η διατήρηση του περιεχομένου, ώστε αυτό να μπορεί να χρησιμοποιηθεί αποτελεσματικά από τις μελλοντικές γενιές.

Το ερωτήματα που θα πρέπει να απαντήσει ο φορέας σε αυτές τις περιπτώσεις είναι τα εξής:

1. Ποια είναι η **μορφή** του πρωτότυπου τεκμηρίου (έγγραφο, φωτογραφία, φιλμ, τρισδιάστατο αντικείμενο, ηχητικό τεκμήριο, βίντεο κλπ)
2. Σε ποια **κατηγορία** ανήκει; (πρωτότυπο αντικείμενο που χρειάζεται ψηφιοποίηση, εξ' αρχής ψηφιακά αρχεία (digital born objects), ψηφιακό αντίγραφο (backup copy))
3. Ποια είναι η **κατάσταση** του πρωτότυπου μη ψηφιακού αντικειμένου; (αντέχει τη διαδικασία ψηφιοποίησης; Έχει κάποιες ιδιαίτερες απαιτήσεις;)
4. Τι **εξοπλισμό** μπορώ να διαθέσω; (εξοπλισμός για την ψηφιοποίηση, μέγεθος σκληρών δίσκων για την αποθήκευση, τεχνολογίες για τη διάθεση κλπ)
5. Υπάρχουν **περιορισμοί**; Αν ναι, ποιοι; (π.χ. ένα βίντεο από παλιές βιντεοκασέτες έχει συγκεκριμένα όρια ποιότητας)
6. Ποιος είναι ο **στόχος**; (πχ. Θέλω να δημιουργήσω ένα διαφημιστικό φυλλάδιο για το μουσείο και χρειάζομαι φωτογραφίες των εκθεμάτων; Θέλω να βιντεοσκοπήσω ένα συνέδριο και να διαθέσω τις ομιλίες στο διαδίκτυο; κλπ.)
7. Ποιο είναι το **χρονοδιάγραμμα**;
8. Πόσα είναι τα **τεκμήρια**;
9. Ποια είναι τα προτεινόμενα **πρότυπα αρχείων**, την χρονική στιγμή της ψηφιοποίησης;

Για να εξασφαλιστεί, η κατά το δυνατότερο μακροχρόνια διαφύλαξη του ψηφιακού αρχείου, πρέπει να ακολουθηθούν κάποιες προδιαγραφές τόσο κατά τη διαδικασία ψηφιοποίησης όσο και κατά την εξ' αρχής δημιουργία ψηφιακών αρχείων. **Οι προδιαγραφές ψηφιοποίησης και ψηφιακών αρχείων παρουσιάζονται αναλυτικά στον παρόν οδηγό του ΕΚΤ.**

1.2 Σύστημα Διαχείρισης και Διάθεσης Περιεχομένου

Τα συστήματα Διαχείρισης και Διάθεσης Περιεχομένου είναι ο πυρήνας της διαδικασίας διάθεσης περιεχομένου προς τους χρήστες και άλλα συστήματα. Φιλοξενούν τα ψηφιακά αρχεία (κείμενο, εικόνα, βίντεο κ.α.) καθώς και τις πληροφορίες σχετικά με την προέλευση, τη φύση, την κατάσταση και την ερμηνεία των αντικειμένων με συστηματικό τρόπο, δηλαδή τα μεταδεδομένα. Περιλαμβάνουν πληροφορίες σχετικά με τις άδειες διάθεσης τόσο των μεταδεδομένων αλλά και των ψηφιακών αρχείων. Παρέχουν στους χρήστες λειτουργίες πλοήγησης, αναζήτησης και προβολής των τεκμηρίων. Σε μερικές περιπτώσεις, μπορεί να περιλαμβάνουν και εξειδικευμένες λειτουργίες προς το προσωπικό του φορέα, όπως λειτουργίες δανεισμού για τις βιβλιοθήκες, οργάνωση εκθέσεων για τα μουσεία κλπ. **Τέλος, υποστηρίζουν τη διαλειτουργικότητα με άλλα συστήματα ενισχύοντας την ανταλλαγή, αναζήτηση, συσσώρευση και διασύνδεση του περιεχομένου με οργανωμένο τρόπο.**

Ως συστήματα Διαχείρισης και Διάθεσης Περιεχομένου μπορούν να οριστούν τα παρακάτω:

1. **αποθετήρια (repositories)**, 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 (digital born).

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

- i. Αποθετήρια / συλλογές ψηφιακού πολιτιστικού ή/ και ιστορικού αποθέματος (π.χ. εικόνες έργων τέχνης, αρχαιολογικών μνημείων, χειρογράφων).
 - ii. Αποθετήρια επιστημονικού περιεχομένου (π.χ. επιστημονικές δημοσιεύσεις όπως άρθρα, βιβλία, πρακτικά συνεδρίων).
 - iii. Αρχεία διάφορων τύπων και μορφών (π.χ. προσωπικά αρχεία προσωπικοτήτων, οπτικοακουστικά αρχεία ραδιοτηλεοπτικών οργανισμών, αρχεία περιοδικών εκδόσεων).
2. **συστήματα διαχείρισης βιβλιοθήκης**, υποστηρίζουν όλες τις απαραίτητες λειτουργικότητες για την ανάπτυξη και διαχείριση: α) μεταδεδομένων, δεδομένων και ψηφιακού υλικού, β) μελών, πράξεων δανεισμού & προσωποποιημένων υπηρεσιών προς τον χρήστη, γ) στατιστικών δεδομένων που προκύπτουν από το σύνολο των δεδομένων της βιβλιοθήκης. Παράλληλα, διαθέτουν πρωτόκολλα για τη διάθεση του περιεχομένου σε άλλα υπολογιστικά συστήματα.
 3. **συστήματα ηλεκτρονικής έκδοσης**, δηλαδή, συστήματα που απευθύνονται σε εκδότες παρέχοντας ψηφιακές υπηρεσίες που καλύπτουν όλα τα στάδια της εκδοτικής δραστηριότητας και εξασφαλίζουν την παρουσίαση και τη διάθεση των ηλεκτρονικών εκδόσεων σύμφωνα με τις τρέχουσες τάσεις και τα διεθνώς αναγνωρισμένα πρότυπα.
 4. **συστήματα διαχείρισης μουσείων**, ολοκληρωμένα σύστημα ειδικά σχεδιασμένο για τις ποικίλες ανάγκες μουσείων και συλλογών: από την επιστημονική τεκμηρίωση ως την οργάνωση μιας έκθεσης, από τα θέματα της συντήρησης ως τις δημόσιες σχέσεις, από την αρχειοθέτηση εικονογραφικού υλικού ως τη διοργάνωση εκδηλώσεων. Τα σύγχρονα συστήματα διαχείρισης μουσείων ενσωματώνουν και λειτουργίες για τη δημοσίευση και τη διάθεση του περιεχομένου στο διαδίκτυο.

1.3 Τεκμηρίωση και μεταδεδομένα

Συχνά ο όρος μεταδεδομένα εξηγείται ως "δεδομένα που περιγράφουν άλλα δεδομένα". Τα μεταδεδομένα είναι δομημένη πληροφορία η οποία προκύπτει από τη διαδικασία της τεκμηρίωσης και η οποία χρησιμοποιείται για την περιγραφή, την επεξήγηση, τον εντοπισμό, τη χρήση και τη διαχείριση μίας πληροφοριακής πηγής, όπως ένα βιβλίο, ένα μουσειακό αντικείμενο ή ένας φάκελος αρχείου. [19]

Τα μεταδεδομένα μπορούν να ταξινομηθούν σε τρεις κατηγορίες ¹:

1. **τα περιγραφικά μεταδεδομένα** χρησιμοποιούνται για την περιγραφή μίας πηγής ώστε να εξασφαλίζεται η περιγραφή, η ταυτοποίηση και η εύρεσή της. Τέτοια στοιχεία αφορούν στον τίτλο, το δημιουργό, τις λέξεις-κλειδιά και τη σύνοψη,

¹ Πηγή: **NISO**

2. **τα δομικά μεταδεδομένα** προσδιορίζουν τον τρόπο με τον οποίο είναι οργανωμένα και δομούνται σύνθετα αντικείμενα, π.χ. πώς οργανώνονται οι σελίδες ενός βιβλίου σε κεφάλαια, και,

3. **τα διαχειριστικά μεταδεδομένα** προσφέρουν πληροφορία που χρησιμοποιείται στη διαχείριση της πηγής, όπως, π.χ. πότε και πώς δημιουργήθηκε η πηγή, το είδος του αρχείου και ποιός μπορεί να έχει πρόσβαση σε αυτή.

1.4 Ανοικτά Διασυνδεδεμένα δεδομένα

Τα διασυνδεδεμένα δεδομένα, χρησιμοποιούν τις κλασσικές τεχνολογίες του παγκοσμίου ιστού, αλλά αντί να διαθέτουν απλές html σελίδες στον χρήστη, επεκτείνουν την πληροφορία έτσι ώστε να δημιουργούνται σημασιολογικές συνδέσεις μεταξύ διαφορετικών συνόλων, οι οποίες είναι αντιληπτές τόσο στον χρήστη αλλά και στις υπολογιστικές μηχανές [19]. Οι συνδέσεις αυτές είναι άπειρες και, μπορούν να οδηγήσουν σε επανάχρηση της γνώσης με τρόπο που δεν μπορούσαμε να το φανταστούμε. *‘It is the unexpected re-use of information which is the value added by the web’ Tim Berners Lee, inventor of the World Wide Web’.*

Τα ανοικτά διασυνδεδεμένα δεδομένα, είναι τα διασυνδεδεμένα δεδομένα, τα οποία δίνονται ανοικτή άδεια που επιτρέπει την επανάχρηση της πληροφορίας, κυρίως τη CC0 ή τη CC BY.

1.5 Συστήματα συσσώρευσης περιεχομένου

Οι συσσωρευτές (Aggregators) συσσωρεύουν μεταδεδομένα από τα επιμέρους συστήματα διαχείρισης και διάθεσης περιεχομένου των φορέων και το διαθέτουν στο ευρύ κοινό μέσω ενός σύγχρονου και φιλικού περιβάλλοντος εμπλουτισμένο με προηγμένες λειτουργίες ενιαίας αναζήτησης και εξελιγμένες δυνατότητες πλοήγησης, εξασφαλίζοντας τη πρόσβαση σε ψηφιακό ή ψηφιοποιημένο περιεχόμενο, ποικίλων τύπων. Παράλληλα, τροφοδοτούν άλλα συστήματα διάθεσης και επανάχρησης του περιεχομένου (π.χ. εκπαιδευτικές εφαρμογές), συμβάλλοντας στην ανάδειξη και την εξάπλωση του περιεχομένου και την ενίσχυση του ρόλου των φορέων.

Στη συνέχεια, δίνονται κάποια συστήματα συσσώρευσης περιεχομένου σε ευρωπαϊκό και εθνικό επίπεδο:

- **Europeana:** Ευρωπαϊκή πύλη συσσώρευσης πολιτιστικού και ιστορικού υλικού
<http://www.europeana.eu/portal/>
- **OpenAire:** Συσσωρεύει δημοσιεύσεις που έχουν προκύψει από χρηματοδοτήσεις από προγράμματα H2020, FP7 & SC39 ή ER Ευρωπαϊκή πύλη συσσώρευσης πολιτιστικού και ιστορικού υλικού
<https://www.openaire.eu/>
- **DART Europe:** Ευρωπαϊκή πύλη συσσώρευσης διδακτορικών διατριβών
<http://www.dart-europe.eu/>
- **SearchCulture:** Εθνικός συσσωρευτής του ΕΚΤ για πολιτιστικό και ιστορικό υλικό
<http://searchculture.gr/>
- **OpenArchives:** Εθνικός συσσωρευτής του ΕΚΤ για ανοικτό περιεχόμενο
<http://www.openarchives.gr/>
- **Μήτιδα:** Διαδικτυακό εργαλείο του για τη δημιουργία ψηφιακού διαδραστικού εκπαιδευτικού περιεχομένου χρησιμοποιώντας συσσωρευμένο υλικό από πηγές περιεχομένου του ΕΚΤ
<http://www.mitida.gr/>

1.6 Επίπεδα διαλειτουργικότητας

Βασικός παράγοντας για να είναι επιτυχής η διαδικασία συλλογής και συσσώρευσης είναι τα **συστήματα διαχείρισης και διάθεσης περιεχομένου να υποστηρίζουν τα απαραίτητα επίπεδα διαλειτουργικότητας**. Ειδικότερα, η διαλειτουργικότητα στο πεδίο των μεταδεδομένων μπορεί να θεωρηθεί πως έχει τις παρακάτω τρεις διαστάσεις [6]:

1) Διαλειτουργικότητα σε επίπεδο συστημάτων. Επιτρέπει την online

διάθεση των μεταδεδομένων ενός αποθετηρίου σε εφαρμογές και συστήματα τρίτων. 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 μετά-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.

2) 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, ESE, 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.

3) 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.

Ψηφιακά αρχεία



Διαλειτουργικότητα

1^ο

σε επίπεδο συστημάτων

2^ο

σε επίπεδο σύνταξης και δομής

3^ο

σε σημασιολογικό επίπεδο

Διάθεση προς τον Χρήστη



Εικόνα 2: Δομή Οδηγού

Στη συνέχεια αυτού του οδηγού, παρουσιάζονται οι **Προδιαγραφές Ψηφιοποίησης και Ψηφιακών Αρχείων**, οι οποίες διασφαλίζουν τη μακροχρόνια διαφύλαξη των ψηφιακών αρχείων και τη διαθεσιμότητα τους σε όσο το δυνατόν περισσότερους χρήστες.

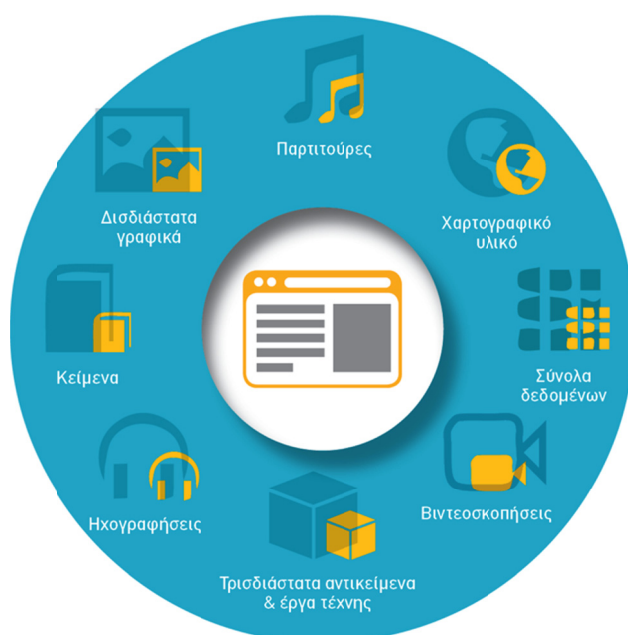
Οι Προδιαγραφές Ψηφιοποίησης και Ψηφιακών Αρχείων, συμπληρώνονται με ένα ακόμη οδηγό, στον οποίο περιγράφονται οι **Προδιαγραφές Διαλειτουργικότητας Ανοικτού Περιεχομένου** με άλλα συστήματα και οι **Προδιαγραφές Διάθεσης προς τον Χρήστη**.

2 Προδιαγραφές για ψηφιακά αρχεία

2.1 Τύποι περιεχομένου ΕΚΤ και ομάδες μορφότυπων αρχείων

“Χρυσή τομή: Προτυποποίηση και ευχρηστία”

Μια από τις πολλές καινοτομίες των αποθετηρίων SaaS του ΕΚΤ είναι η ισορροπία μεταξύ προτυποποίησης και ευχρηστίας. Γι’ αυτό το λόγο, στα πλαίσια του έργου SaaS για τα αποθετήρια, το ΕΚΤ έχει ορίσει 8 τύπους περιεχομένου. Οι τύποι αυτοί καλύπτουν κάθε είδος υλικού που μπορεί να συλλέγει ένας φορέας. Οι τύποι περιεχομένου ορίζονται από ένα σύνολο μεταδεδομένων προσαρμοσμένο στις απαιτήσεις του κάθε τύπου και τα ψηφιακά αρχεία.



Τα ψηφιακά αρχεία μπορεί να είναι διαφορετικών μορφότυπων και κάποιοι μορφότυποι μπορεί να επαναλαμβάνονται σε διαφορετικούς τύπους περιεχομένου. Για παράδειγμα, μια σελίδα ψηφιοποιημένου βιβλίου (δηλ. Τύπος περιεχομένου: Κείμενο), μπορεί να περιλαμβάνει ως ψηφιακά αρχεία, ένα αρχείο κειμένου pdf και την αρχικά ψηφιοποιημένη εικόνα σε μορφή tiff. Αντίστοιχα, ένα έργο τέχνης (δηλ. Τύπος περιεχομένου: Τρισδιάστατα αντικείμενα & έργα τέχνης) μπορεί να έχει ως ψηφιακό αρχείο μια εικόνα tiff. Γι’ αυτό το σκοπό, έχουν οριστεί συγκεκριμένες

ομάδες μορφότυπων περιεχομένου [7].

Συγκεκριμένα:

1. Αρχεία εικόνας
2. Αρχεία κειμένου
3. Αρχεία Βίντεο
4. Ηχητικά αρχεία
5. Σύνολα δεδομένων

Στη συνέχεια παρουσιάζουμε τις προδιαγραφές για κάθε ομάδα μορφότυπου ξεχωριστά.



2.2 Προδιαγραφές για αρχεία εικόνας

Πρόκειται για ψηφιακό υλικό που αναπαριστά είτε δισδιάστατα αντικείμενα, όπως ζωγραφικά έργα, φωτογραφίες, καρτ-ποστάλ, σχέδια κ.α. είτε τρισδιάστατα αντικείμενα όπως έργα γλυπτικής, εκθέματα, παιχνίδια κ.α. Επιπλέον, τα αρχεία εικόνας μπορούν να παραχθούν εξ' αρχής με τη χρήση ψηφιακών μέσων και προγραμμάτων υπολογιστή.

Συνεπώς, τα αρχεία εικόνας μπορεί να προκύψουν από τρεις διαφορετικούς τρόπους, τους οποίους εξετάζουμε αναλυτικά στη συνέχεια:

1) Ψηφιοποίηση δισδιάστατων αντικειμένων (2D): Φωτογραφίες ή αρνητικά φωτογραφιών, σχέδια, ζωγραφικά έργα, κάρτες δραστηριοτήτων, σχεδιαγράμματα, κολάζ, καρτ-ποστάλ, αφίσες, εκτυπώσεις, τεχνικά σχέδια, κ.α.

2) Ψηφιοποίηση (Φωτογράφιση) τρισδιάστατων αντικειμένων (3D): Πρόκειται για αντικείμενα που αποτελούνται από τρεις διαστάσεις, δηλαδή από μήκος, πλάτος και ύψος, π.χ. γλυπτά, νομίσματα και άλλα τρισδιάστατα έργα τέχνης ή τις απομιμήσεις τους, εκθέματα, μηχανές, είδη ένδυσης, παιχνίδια, και εργόχειρα, φυσικά αντικείμενα όπως δείγματα μικροσκοπίου ή άλλα δείγματα που είναι τοποθετημένα για προβολή π.χ. σε μουσεία ή εκθέσεις

3) Εξ' αρχής παραγωγή ψηφιακών αρχείων εικόνας (digital born): Πρόκειται για αρχεία εικόνων τα οποία παράγονται εξ' αρχής με τη χρήση εξειδικευμένων προγραμμάτων υπολογιστή, πχ. Adobe Photoshop, Adobe Illustrator, AutoCAD κλπ.

Σε πολλές περιπτώσεις, οι εικόνες μπορεί να περιέχουν ή να έχουν συνοδευτικό γραπτού υλικού. Λόγω αυτής της διπλής φύσης, η ψηφιοποίηση των εικόνων είναι παρόμοια με την ψηφιοποίηση του κειμένου, την οποία αναλύουμε στην επόμενη ενότητα.

Για τις τρεις παραπάνω περιπτώσεις, η διαδικασία θα πρέπει να οδηγήσει στην παραγωγή τριών τύπων εικόνων [12][13]:

Πρωτότυπη εικόνα	Εικόνα διάθεσης στο διαδίκτυο	Εικόνα προεπισκόπησης (thumbnail)
1. Αναπαριστά με όσο το δυνατόν περισσότερη πληροφορία την αρχική εικόνα 2. Δεν έχει υποβληθεί σε συμπίεση ή η συμπίεση έχει γίνει χωρίς απώλειες 3. Δεν έχει υποβληθεί σε τροποποιήσεις 4. Χρησιμοποιείται για την μακροπρόθεσμη διαφύλαξη, για την παραγωγή παράγωγων αρχείων και την εκτύπωση 5. Είναι υψηλής ποιότητας 6. Αποθηκεύεται κυρίως σε μορφή αρχείου TIFF ή JPEG2000 7. Διατηρείται και η αρχική μορφή του αρχείου αν είναι διαφορετική από TIFF ή JPEG2000	1. Χρησιμοποιείται στη θέση της πρωτότυπης ψηφιοποιημένης εικόνας για διάθεση στο διαδίκτυο 2. Πρέπει να χωράει μέσα στη περιοχή προβολής της μέσης οθόνης υπολογιστή 3. Λογικό μέγεθος αρχείου για γρήγορο χρόνο λήψης, χωρίς να απαιτείται γρήγορη σύνδεση στο διαδίκτυο 4. Αποδεκτή ποιότητα για γενική έρευνα και χρήση 5. Συνήθως είναι συμπίεσμνης μορφής 6. Συνήθως αποθηκεύονται σε μορφή JPEG ή μορφή αρχείου JPEG2000	1. Σχεδιασμένη για την γρήγορη εμφάνιση στο διαδίκτυο 2. Θα πρέπει να αντιπροσωπεύει την πρωτότυπη ψηφιοποιημένη εικόνα ώστε να επιτρέπει στον χρήστη να αποφασίσει αν θέλει να δει την εικόνα διάθεσης 3. Συνήθως αποθηκεύονται σε μορφή GIF ή JPEG μορφές 4. Δεν είναι πάντα κατάλληλη για εικόνες που αποτελούνται κυρίως από κείμενο, παρτιτούρες, κλπ . Σε αυτές τις περιπτώσεις, ο χρήστης δεν μπορεί να διακρίνει το περιεχόμενο σε τόσο μικρή κλίμακα

Πίνακας 1: Τύποι αρχείων εικόνας

2.2.1 Ψηφιοποίηση δισδιάστατων αντικειμένων (2D)

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 [2][3][4][5].

Αναφορικά με την πρωτότυπη ψηφιοποιημένη εικόνα, ιδιαίτερη σημασία έχουν οι ελάχιστες απαιτήσεις ποιότητας που αναφέρονται στην μελέτη [1], οι οποίες αναπαράγονται στη συνέχεια, αναφορικά με τα δισδιάστατα αντικείμενα:

Πρωτότυπο αντικείμενο	Ελάχιστη ανάλυση <i>Σημείωση: Το dpi αναφέρεται σε σχέση με το φυσικό μέγεθος του αντικειμένου</i>	Χρωματικό βάθος	Προτεινόμενος Μορφότυπος
Φωτοτυπημένο υλικό (ασπρόμαυρο)	200-300 dpi	8 bit γκρι	TIFF ή JPEG 2000
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	
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 2. Ελάχιστες απαιτήσεις για ψηφιοποίηση δισδιάστατων 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.).

Οι απαιτήσεις για την **εικόνα διάθεσης** και την **εικόνα προεπισκόπησης**, συνοψίζονται στον παρακάτω πίνακα (Πίνακας 3). [9]

	Ελάχιστη ανάλυση	Χρωματικό βάθος	Προτεινόμενος Μορφότυπος
Εικόνα διάθεσης στο διαδίκτυο	72-300 ppi και 800-3000 pixels στη μεγαλύτερη διάσταση	8 bit (ασπρόμαυρη) ή 24 bit (έγχρωμη)	JPG ή JPEG2000
Εικόνα προεπισκόπησης (thumbnail)	72 ppi και 150-200 pixels στη μεγαλύτερη διάσταση	8 bit (ασπρόμαυρη) ή 24 bit (έγχρωμη)	

Πίνακας 3: Προδιαγραφές για την Εικόνα Διάθεσης και την Εικόνα Προεπισκόπησης διςδιάστατων αντικειμένων (2d) [9]



Διαβάστε επίσης...

- ➔ Συγκριτική μελέτη μορφότυπων εικόνας (TIFF, PNG, JPEG2000, JPG, PDF) που χρησιμοποιούνται στην ψηφιοποίηση από το Federal Agencies Digitization Guidelines Initiative – FADGI (September 2, 2014) http://www.digitizationguidelines.gov/guidelines/FADGI_RasterFormatCompare_p1_20140902_r.pdf

2.2.2 Ψηφιοποίηση τρισδιάστατων αντικειμένων (3D)

Η ψηφιοποίηση των τρισδιάστατων αντικειμένων, αφορά τη διαδικασία αναπαράστασης του τρισδιάστατου αντικειμένου σε διςδιάστατη εικόνα, και συνήθως γίνεται με τη χρήση ψηφιακής φωτογραφικής μηχανής DSLR. Η μελέτη 'Guidelines for the Creation of Digital Collections: Digitization Best Practices for Three-Dimensional Objects' [10] περιλαμβάνει αναλυτικές οδηγίες για την φωτογράφιση ενός τρισδιάστατου αντικειμένου, αναφορικά με τον εξοπλισμό που χρειάζεται, τη διαδικασία φωτογράφισης, τις ρυθμίσεις της φωτογραφικής μηχανής και της προδιαγραφές του τελικού αρχείου που προκύπτει. Στον παρακάτω πίνακα (Πίνακας 4) συνοψίζουμε τις ρυθμίσεις της μηχανής όπως παρουσιάζονται στο [10] καθώς και τις απαιτήσεις ανάλυσης και βάθους χρώματος του παραγόμενου αρχείου εικόνας, το οποίο αφορά την **πρωτότυπη εικόνα**.

Είδος Ρύθμισης	Προτεινόμενες τιμές	Σχόλια
ISO	μεταξύ 100-200	Ένα χαμηλότερο ISO παράγει μια ομαλότερη εικόνα με λιγότερο ψηφιακό «θόρυβο»
Ταχύτητα κλείστρου	1/25	Αυτό μπορεί να απαιτεί κάποιο πειραματισμό, ανάλογα με τη ρύθμιση που χρησιμοποιείται. Αν οι φωτογραφίες είναι πολύ σκοτεινές, μπορεί να απαιτηθεί μεγαλύτερη ταχύτητα κλείστρου. Ωστόσο, μια μεγαλύτερη ταχύτητα κλείστρου μπορεί να προκαλέσει θόλωση της εικόνας, ειδικά εάν δεν χρησιμοποιείται τρίποδο.

Μέγεθος διαφράγματος	f8 - f11	Το μικρότερο διάφραγμα αναπαριστά καλύτερα το βάθος πεδίου στην εικόνα (αντικείμενα τόσο κοντά όσο και μακριά βρίσκονται στο επίκεντρο (focus)).
Ισορροπία λευκού	Αυτόματη	Αν γίνεται αποθήκευση των εικόνων σε μορφή RAW, η ισορροπία λευκού μπορεί να ρυθμιστεί αργότερα
Ρυθμίσεις αρχείου	>10 - 16 megapixels, 48-bit RGB δηλ. 300 ppi και >5000 pixels στη μεγαλύτερη διάσταση (5760x3840)	Οι αρχικές εικόνες πρέπει να λαμβάνονται στην υψηλότερη ανάλυση/ποιότητα που παρέχει η ψηφιακή κάμερα.
Είδος αρχείου	RAW (για την φωτογράφιση) TIFF (για τη διαφύλαξη)	Τα αρχεία RAW αποθηκεύουν την εικόνα όπως ακριβώς τη 'βλέπει' ο αισθητήρας της κάμερας και επιτρέπει την εύκολη προσαρμογή των ρυθμίσεων χρώματος και ισορροπίας λευκού. Συστήνεται να χρησιμοποιείται κατά τη διάρκεια της φωτογράφισης. Επειδή η κωδικοποίηση του RAW αρχείου αλλάζει ανάλογα με τον κατασκευαστή της κάμερας, συστήνεται η αποθήκευση της φωτογραφίας σε μορφή TIFF, χωρίς συμπίεση, για τη μακροχρόνια διαφύλαξη του αρχείου.

Πίνακας 4: Προδιαγραφές ψηφιοποίησης με φωτογράφιση [29]

Οι απαιτήσεις για την **εικόνα διάθεσης** και την **εικόνα προεπισκόπησης** είναι οι ίδιες με τις αντίστοιχες προδιαγραφές για δισδιάστατα αντικείμενα (2d) και συνοψίζονται στον παρακάτω πίνακα (Πίνακας 5) [15].

	Ελάχιστη ανάλυση	Χρωματικό βάθος	Προτεινόμενος Μορφότυπος
Εικόνα διάθεσης στο διαδίκτυο	72-300 ppi και 800-3000 pixels στη μεγαλύτερη διάσταση	8 bit (ασπρόμαυρη) ή 24 bit (έγχρωμη)	JPG ή JPEG2000
Εικόνα προεπισκόπησης (thumbnail)	72 ppi και 150-200 pixels στη μεγαλύτερη διάσταση	8 bit (ασπρόμαυρη) ή 24 bit (έγχρωμη)	

Πίνακας 5: Προδιαγραφές για την Εικόνα Διάθεσης και την Εικόνα Προεπισκόπησης τρισδιάστατων αντικειμένων (3d)

Σημείωση: Σε αρκετές περιπτώσεις, η αναπαράσταση του τρισδιάστατου αντικειμένου μπορεί να γίνει και με ειδικά προγράμματα παραγωγής τρισδιάστατων απεικονίσεων (π.χ. Autodesk 123D Catch). Καθώς δεν υπάρχουν ακόμα σαφείς οδηγίες από την παγκόσμια κοινότητα, τα αρχεία που προκύπτουν με αυτόν τον τρόπο δεν αποτελούν μέρος της συγκεκριμένης μελέτης.



Διαβάστε επίσης...

- ➔ Πως πρέπει να μεταχειριστούμε τα πρωτότυπα αντικείμενα κατά την ψηφιοποίηση (Ενότητα 2.5 από τον «Οδηγό 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
http://digitization.hpclab.ceid.upatras.gr/Odhgos_kalwn_praktikwn1.0.pdf
- ➔ Συγκριτική μελέτη μορφότυπων εικόνας (TIFF, PNG, JPG2000, JPG, PDF) που χρησιμοποιούνται στην ψηφιοποίηση από το Federal Agencies Digitization Guidelines Initiative – FADGI (September 2, 2014)
http://www.digitizationguidelines.gov/guidelines/FADGI_RasterFormatCompare_p1_20140902_r.pdf

2.2.3 Εξ' αρχής παραγωγή ψηφιακών αρχείων εικόνας (digital born)

Πρόκειται για αρχεία εικόνων τα οποία παράγονται εξ' αρχής με τη χρήση εξειδικευμένων προγραμμάτων υπολογιστή, πχ. Adobe Photoshop, Adobe Illustrator, AutoCAD κλπ. Σε αυτές τις περιπτώσεις, ανάλογα με το πρόγραμμα που χρησιμοποιείται, υπάρχουν χιλιάδες είδη αρχείων και τρόποι κωδικοποίησης που μπορεί να προκύψουν. Σε αυτές τις περιπτώσεις συστήνονται τα παρακάτω:

- 1) Πριν από τη δημιουργία του αρχείου, θα πρέπει να μελετηθούν τα προτεινόμενα ή/και πιο δημοφιλή προγράμματα και μορφότυποι αρχείων, που μπορούν να χρησιμοποιηθούν ανάλογα με τις απαιτήσεις και τους στόχους (πχ. για ένα αρχιτεκτονικό σχέδιο, συστήνεται να χρησιμοποιηθεί το AutoCAD, το οποίο παράγει αρχεία AutoCAD .dwg file format)
- 2) Αποθήκευση του ψηφιακού αρχείου εικόνας στο προτεινόμενο (πρωτεύων) μορφότυπο ανάλογα με το πρόγραμμα που χρησιμοποιείται (π.χ. dwg για το πρόγραμμα AutoCAD)
- 3) Μετατροπή του αρχείου και αποθήκευση αντίγραφου σε ένα πιο δημοφιλές μορφότυπο, ο οποίος δεν εξαρτάται από το πρόγραμμα που χρησιμοποιήθηκε για την παραγωγή του. Με αυτόν τον τρόπο, εξασφαλίζεται η μακροχρόνια διαφύλαξη του αρχείου, τουλάχιστον για προβολή. Η βιβλιοθήκη του Κογκρέσου [11] και η Ιστορική Βιβλιοθήκη Bentley του Πανεπιστημίου του Michigan [12], προτείνουν την μετατροπή προβληματικών αρχείων σε συγκεκριμένους μορφότυπους (Πίνακας 6).

Είδος αρχείου	Αρχικός μορφότυπος	Προτεινόμενος μορφότυπος για ασφαλή διαφύλαξη
Raster Εικόνα / Raster Image Files	RAW	
	BMP: Windows Bitmap	
	PSD: Adobe Photoshop Document	
	FPX: FlashPix Bitmap	TIFF (Uncompressed Bitmap)
	PCD: Kodak Photo CD Image	ή JPEG 2002 (jp2)
	PCT: Apple Picture File	
	TGA: Targa Graphic	

**Διανυσματικά
γραφικά /
Vector Image Files**

AI: Adobe Illustrator	
CDR: CorelDraw	
CMX: Corel Exchange (CMX)	SVG Format
DRW : Micrografx Draw	
WMF: Windows Metafile	
PS: PostScript	
EPS: Encapsulated PostScript	TIFF (Uncompressed Bitmap) ή PDF/A Format

Πίνακας 6: Προτεινόμενοι μορφότυποι και μετατροπές για εξ' αρχής ψηφιακά αρχεία εικόνας

Οι απαιτήσεις για την **εικόνα διάθεσης** και την **εικόνα προεπισκόπησης** είναι οι ίδιες με τις αντίστοιχες προδιαγραφές για δισδιάστατα αντικείμενα (2d) και συνοψίζονται στον παρακάτω πίνακα (Πίνακας 7) [15].

	Ελάχιστη ανάλυση	Χρωματικό βάθος	Προτεινόμενος Μορφότυπος
Εικόνα διάθεσης στο διαδίκτυο	72-300 ppi και 800-3000 pixels στη μεγαλύτερη διάσταση	8 bit (ασπρόμαυρη) ή 24 bit (έγχρωμη)	JPG ή JPEG2000
Εικόνα προεπισκόπησης (thumbnail)	72 ppi και 150-200 pixels στη μεγαλύτερη διάσταση	8 bit (ασπρόμαυρη) ή 24 bit (έγχρωμη)	

Πίνακας 7: Προδιαγραφές για την Εικόνα Διάθεσης και την Εικόνα Προεπισκόπησης εξ' αρχής ψηφιακών εικόνων

Για οποιονδήποτε κι από τους τρεις τρόπος που μπορεί να έχει δημιουργηθεί το αρχείο εικόνας, συστήνεται η χρήση τεχνικών μεταδεδομένων για να εξασφαλιστεί η μακροχρόνια διαφύλαξη του αρχείου [20]. Για το σκοπό αυτό μπορεί να χρησιμοποιηθεί είτε το σχήμα μεταδεδομένων PREMIS [21] ή METS [22].

</> Τεχνικά μεταδεδομένα

- ➔ μορφότυπος αρχείου
- ➔ ανάλυση εικόνας (σε ppi)
- ➔ διαστάσεις εικόνας (σε cm ή inch)
- ➔ χρωματικό βάθος
- ➔ κωδικοποίηση χρώματος (π.χ. RGB, CMYK, grayscale)
- ➔ στοιχεία σκάνερ ή ψηφιακής κάμερας που χρησιμοποιήθηκε: μάρκα, όνομα, αριθμός
- ➔ το λογισμικό που χρησιμοποιήθηκε για τη διαχείριση της εικόνας (όνομα προγράμματος και έκδοση)



2.3 Προδιαγραφές για αρχεία κειμένου

Πρόκειται για τεκμήρια που στο μεγαλύτερο μέρος τους αποτελούνται από λέξεις ή σύμβολα που διαβάζονται σαν λέξεις και προτάσεις.

Περιλαμβάνει όλες τις μορφές της γλώσσας και γραφής συμπεριλαμβανομένων και εκείνων που γίνονται αντιληπτές μέσω της αφής π.χ. γραφή braille σε συμβατικά ή ψηφιακά μέσα.

Συνεπώς τα αρχεία κειμένου περιλαμβάνουν έντυπο υλικό όπως βιβλία, περιοδικά, και εφημερίδες, ή χειρόγραφα ή δακτυλογραφημένα πρωτότυπα κείμενα, επιστολές, σημειώσεις ή ψηφιακά κείμενα που έχουν πληκτρολογηθεί στον υπολογιστή. Για τους σκοπούς του παρόντος εγγράφου, θα διαχωρίσουμε τους τύπους των αρχείων σε:

1) Αρχεία κειμένου προς ψηφιοποίηση: Πρόκειται για αντικείμενα που έχουν ήδη αποτυπωθεί σε ένα φυσικό υλικό. Μπορεί να περιλαμβάνει α) έντυπο υλικό όπως βιβλία, περιοδικά, και εφημερίδες, ή β) χειρόγραφα όπως χειρόγραφα βιβλία, σημειώσεις και επιστολές ή γ) δακτυλογραφημένο υλικό όπως πρωτότυπα κείμενα, επιστολές, σημειώσεις κλπ.

2) Εξ' αρχής παραγωγή ψηφιακών αρχείων κειμένου (digital born): Πρόκειται για αρχεία κειμένου τα οποία παράγονται εξ' αρχής με τη χρήση εξειδικευμένων προγραμμάτων υπολογιστή, πχ. Microsoft Office, Open Office, QuarkXPress Adobe Illustrator, κλπ.

Για τις δυο παραπάνω περιπτώσεις, η διαδικασία θα πρέπει να οδηγήσει στην παραγωγή τριών τύπων αρχείων [13][14]:

Πρωτότυπα αρχεία κειμένου	Κείμενο διάθεσης στο διαδίκτυο	Εικόνα προεπισκόπησης (thumbnail)
1. Αναπαριστά με όσο το δυνατόν περισσότερη πληροφορία την αρχική εικόνα	1. Χρησιμοποιείται στη θέση του πρωτότυπου αρχείου	1. Σχεδιασμένη για την γρήγορη εμφάνιση στο διαδίκτυο
2. Δεν έχει υποβληθεί σε συμπίεση ή η συμπίεση έχει γίνει χωρίς απώλειες	2. Λογικό μέγεθος αρχείου για γρήγορο χρόνο λήψης, χωρίς να απαιτείται γρήγορη σύνδεση στο διαδίκτυο	2. Θα πρέπει να αντιπροσωπεύει την πρωτότυπη ψηφιοποιημένη εικόνα ώστε να επιτρέπει στον χρήστη να αποφασίσει αν θέλει να δει την εικόνα διάθεσης.
3. Δεν έχει υποβληθεί σε τροποποιήσεις	3. Αποδεκτή ποιότητα για γενική έρευνα και χρήση	3. Δεν πρέπει να συγχέεται με το λογότυπο του
4. Χρησιμοποιείται για την μακροπρόθεσμη διαφύλαξη, για την	4. Συνήθως είναι συμπιεσμένης μορφής	

παραγωγή παράγωγων αρχείων και την εκτύπωση 5. Είναι υψηλής ποιότητας 6. Για ψηφιοποιημένο υλικό, διατηρούνται οι ψηφιοποιημένες εικόνες κυρίως σε μορφή αρχείου TIFF ή JPEG2000. Αν τα αρχεία έχουν περάσει από OCR, διατηρούνται τα αποτελέσματα σε TXT και XML μορφή και παράγεται PDF με τις εικόνες στη μέγιστη ανάλυση με ενσωματωμένο το κείμενο (image pdf with embedded text) 7. Για εξ' αρχής ψηφιακά αρχεία, διατηρείται και η αρχική μορφή του αρχείου.	5. Συνήθως αποθηκεύονται σε μορφή PDF. Σε εξαιρετικές περιπτώσεις μπορεί να δοθεί σε JPEG ή μορφή αρχείου JPEG2000 6. Αν το πρωτότυπο αρχείο έχει προκύψει από ψηφιοποίηση, ενσωματώνει το OCR κείμενο έτσι ώστε να είναι ευρετηριάσιμο από τις μηχανές αναζήτησης	μορφότυπου του αρχείου (π.χ. χρήση λογότυπου της Adobe Acrobat για pdf αρχεία). 4. Συνήθως αποθηκεύονται σε μορφή GIF ή JPEG μορφές 5. Δεν είναι πάντα κατάλληλη για εικόνες που αποτελούνται κυρίως από κείμενο, παρτιτούρες, κλπ . Σε αυτές τις περιπτώσεις, ο χρήστης δεν μπορεί να διακρίνει το περιεχόμενο σε τόσο μικρή κλίμακα. Συστήνεται να επιλέγονται εξώφυλλα.
---	--	---

Πίνακας 8: Τύποι αρχείων κειμένου

2.3.1 Αρχεία κειμένου προς ψηφιοποίηση

Το πρώτο στάδιο της ψηφιοποίησης περιλαμβάνει την παραγωγή της ψηφιακής εικόνας. Γι' αυτό το σκοπό, θα πρέπει να ακολουθηθούν οι διαδικασίες **Ψηφιοποίησης δισδιάστατων αντικειμένων** που περιγράφονται στην παραπάνω ενότητα.

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**. Με αυτό τον τρόπο, είναι δυνατή η ευρετηρίαση (indexing) του κειμένου από μηχανές αναζήτησης, το σύστημα διαχείρισης περιεχομένου του φορέα και συσσωρευτές ανοικτού ψηφιακού περιεχομένου, υποστηρίζοντας την αναζήτηση σε πλήρες κείμενο (full text search).

Προτείνεται να απαιτείται τουλάχιστον «αδιόρθωτο», όπως ονομάζεται, OCR για κείμενα Νέων Ελληνικών² 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

² Αφορά κείμενα από τυπογραφικά δοκίμια και όχι χειρόγραφα κείμενα.

περιεχομένου, εφόσον είναι διαθέσιμα εκεί εκτός των μεταδεδομένων και τα δεδομένα οπτικής αναγνώρισης χαρακτήρων που έχουν προκύψει από τη 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.

Οι παραπάνω απαιτήσεις δίνονται στον κάτωθι πίνακα (Πίνακας 9) σε μορφή κατάλληλη για συμπερίληψη σε πίνακες συμμόρφωσης.

A/A	Απαίτηση	Μορφότυπος	Περιγραφή
1.	Οπτική Αναγνώριση Χαρακτήρων (OCR) για υποστήριξη αναζήτησης και ευρετηρίασης του πλήρους κειμένου	TXT	Παραγωγή αδιόρθωτου OCR για το σύνολο του προς ψηφιοποίηση έντυπου περιεχομένου
2.	Μορφή αποτελεσμάτων OCR	TXT ή XML	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	XML	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

4. Παράδοση αρχείου	τελικού	PDF	αναζήτησης με hit highlighting. Παράδοση αρχείων σε μορφή PDF, ενσωμάτωση στο PDF αρχείο «Image PDF with hidden text» που επιτρέπει την αναζήτηση και επιλογή κειμένου
---------------------	---------	-----	---

Πίνακας 9. Επιπρόσθετες απαιτούμενες προδιαγραφές ψηφιοποίησης

2.3.2 Εξ' αρχής παραγωγή ψηφιακών αρχείων κειμένου (digital born)

Πρόκειται για αρχεία κειμένου τα οποία παράγονται εξ' αρχής με τη χρήση εξειδικευμένων προγραμμάτων υπολογιστή, πχ. Microsoft Office, Open Office, QuarkXPress Adobe Illustrator, κλπ.

- 1) Πριν από την= δημιουργία του αρχείου, θα πρέπει να μελετηθούν τα προτεινόμενα ή/και πιο δημοφιλή προγράμματα και μορφότυποι αρχείων, που μπορούν να χρησιμοποιηθούν ανάλογα με τις απαιτήσεις και τους στόχους. Η βιβλιοθήκη του Κογκρέσου [11] και η Ιστορική Βιβλιοθήκη Bentley του Πανεπιστημίου του Michigan [12], προτείνουν μια σειρά μορφότυπων για αρχεία κειμένου και προτείνουν να είναι ανοικτού τύπου και να βασίζονται σε xml κωδικοποίηση (xml based)

Προτεινόμενοι μορφότυποι...	
για αρχεία κειμένου	DOCX: MS Word Open XML Document (created in MS Office 2007 and 2010) XLSX: MS Excel Open XML Document (created in MS Office 2007 and 2010) PPTX: MS PowerPoint Open XML Document (created in MS Office 2007 and 2010) ODT: OpenDocument Text Document (created in OpenOffice) ODS: OpenDocument Spreadsheet (created in OpenOffice) ODP: OpenDocument Presentation (created in OpenOffice) PDF/A: Portable Document Format (Archival) TXT: Plain Text File (ANSI or UTF-8 encoded) RTF: Rich Text Format File XML: Extensible Markup Language Data File HTML: hierarchy or network of linked pages CSV: Comma Separated Values File TSV: Tab Separated Values File
Αρχεία Email	MBOX Format

Πίνακας 10: Προτεινόμενοι μορφότυποι για εξ' αρχής ψηφιακά αρχεία κειμένου

- 2) Αποθήκευση του ψηφιακού αρχείου κειμένου στο προτεινόμενο (πρωτεύων) μορφότυπο ανάλογα με το πρόγραμμα που χρησιμοποιείται

- 3) Μετατροπή του αρχείου και αποθήκευση αντίγραφου σε ένα πιο δημοφιλές μορφότυπο, ο οποίος δεν εξαρτάται από το πρόγραμμα που χρησιμοποιήθηκε για την παραγωγή του. Με αυτόν τον τρόπο, εξασφαλίζεται η μακροχρόνια διαφύλαξη του αρχείου, τουλάχιστον για προβολή. Η βιβλιοθήκη του Κογκρέσου [11] και η Ιστορική Βιβλιοθήκη Bentley του Πανεπιστημίου του Michigan [12], προτείνουν την μετατροπή αρχείων που μπορεί να μην υποστηρίζονται μελλοντικά σε συγκεκριμένους μορφότυπους (Πίνακας 11).

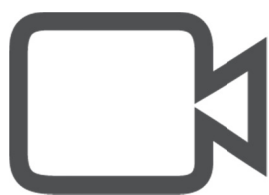
Είδος αρχείου	Αρχικός μορφότυπος	Προτεινόμενος μορφότυπος για ασφαλή διαφύλαξη
Αρχεία Κειμένου	DOC: MS Word 1997-2003 Document	
	PPT: MS PowerPoint 1997-2003 Presentation	MS Office Open XML(OOXML) Format (π.χ. docx, pptx, xlsx κλπ)
	XLS: MS Excel 1997-2003 Spreadsheet	
Αρχεία Email	EML: Email Message File	
	PST: Outlook Personal Information Store File	MBOX Format
	Eudora Mail and approx. 40 other formats	

Πίνακας 11: Προτεινόμενες μετατροπές σε αποδεκτούς μορφότυπους για εξ' αρχής ψηφιακά αρχεία κειμένου

Για οποιονδήποτε κι από τους δύο τρόπος που μπορεί να έχει δημιουργηθεί το αρχείο κειμένου, συστήνεται η χρήση τεχνικών μεταδεδομένων για να εξασφαλιστεί η μακροχρόνια διαφύλαξη του αρχείου [20]. Για το σκοπό αυτό μπορεί να χρησιμοποιηθεί είτε το σχήμα μεταδεδομένων PREMIS [21] ή METS [22].

</> Τεχνικά μεταδεδομένα

- ➔ πληροφορίες για τον τρόπο που παράχθηκε το αρχείο (ποιότητα, πλατφόρμα, το λογισμικό, μέσο)
- ➔ πληροφορίες για τους χαρακτήρες που χρησιμοποιήθηκαν (σύνολο χαρακτήρων και το μέγεθος, το μέγεθος byte, το χαρακτήρα αλλαγής γραμμής)
- ➔ γλώσσες του κειμένου
- ➔ γραμματοσειρές που χρησιμοποιήθηκαν
- ➔ πληροφορίες σχετικά με τη σήμανση και τη διαμόρφωση του κειμένου
- ➔ σημειώσεις σχετικά με την επεξεργασία που ακολουθήθηκε
- ➔ τεχνικές απαιτήσεις για την εκτύπωση και την προβολή
- ➔ σειρά σελίδων



2.4 Προδιαγραφές για αρχεία βίντεο

Πρόκειται για υλικό που χρησιμοποιείται κυρίως για να προβάλλεται με τη συμβολή σχετικών μέσων ή λογισμικού και μπορεί να είναι ταινίες, υλικό βιντεοσκόπησης, καταγεγραμμένες εικόνες σε κίνηση που μπορεί να συνοδεύονται από ήχο, διαφάνειες, slides κ.λπ.

Το υλικό αυτό μπορεί να προκύψει από δύο διαφορετικούς τρόπους:

- 1) Ψηφιοποίηση βίντεο από αναλογικά μέσα (φίλμ, βιντεοκασέτες κλπ):** Πρόκειται για παλιότερο υλικό το οποίο έχει καταγραφεί σε αναλογικά μέσα, όπως μπομπίνες φίλμ, βιντεοκασέτες
- 2) Ψηφιακό βίντεο (digital born):** Πρόκειται για βίντεο το οποίο έχει καταγραφεί χρησιμοποιώντας ψηφιακά μέσα, όπως ψηφιακές κάμερες ή έχει παραχθεί εξ' αρχής με τη χρήση εξειδικευμένων προγραμμάτων υπολογιστή, π.χ. animated film

Και στις δυο περιπτώσεις, οι διαδικασίες θα πρέπει να οδηγήσει στην παραγωγή τριών τύπων αρχείων:

Αρχείο βίντεο για διαφύλαξη	Αρχείο βίντεο για διάθεση στο διαδίκτυο	Εικόνα προεπισκόπησης (thumbnail)
1. Το βίντεο στην καλύτερη ανάλυση 2. Δεν έχει υποβληθεί σε συμπίεση ή η συμπίεση έχει γίνει χωρίς απώλειες 3. Δεν έχει υποβληθεί σε τροποποιήσεις 4. Είναι υψηλής ποιότητας 5. Αποθηκεύεται κυρίως σε μορφή αρχείου mp4 (pure), mpeg ή avi (για	1. Χρησιμοποιείται στη θέση του πρωτότυπου για διάθεση στο διαδίκτυο 2. Πρέπει να χωράει μέσα σε περιοχή προβολής της μέσης οθόνης υπολογιστή 3. Λογικό μέγεθος αρχείου για γρήγορο χρόνο λήψης, χωρίς να απαιτείται γρήγορη σύνδεση στο διαδίκτυο	1. Συστήνεται να παράγεται μια εικόνα προεπισκόπησης, η οποία να αντιπροσωπεύει το βίντεο ώστε να επιτρέπει στον χρήστη να αποφασίσει αν θέλει να το δει (π.χ. εικόνα από το καρέ στους τίτλους μιας ταινίας). Αν δεν υπάρχουν θέματα με πνευματικά δικαιώματα, μπορεί να είναι και η αφίσα της ταινίας. 2. Συνήθως αποθηκεύονται σε

6. Αποθηκεύεται σε ανάλυση 1980x1080 (full HD) ή 720x400 ή 720x576 (για 4:3)	4. Συνήθως είναι συμπίεσμμένης μορφής 5. Συνήθως αποθηκεύονται σε mp4 6. Αποθηκεύεται σε ανάλυση 1280x720 ή 720x400 ή 640x480 (για 4:3)	μορφή GIF ή JPEG μορφές
---	--	--------------------------------

Πίνακας 12: Τύποι αρχείων που προκύπτουν από το βίντεο

Όλες οι μορφές βίντεο (κινηματογραφικά φιλμ, αναλογικά βίντεο, ψηφιακά βίντεο) αποτελούνται από μια ακολουθία εικόνων (καρέ / frames), οι οποίες εμφανίζονται διαδοχικά η μια μετά την άλλη με ένα σταθερό ρυθμό, παράγοντας έτσι την ψευδαίσθηση της κίνησης. Η ακολουθία εικόνας μπορεί επίσης να συνδυάζεται με μια ροή ήχου που αποτελείται από ένα (mono) ή δύο (stereo) κανάλια. Τα ψηφιακά βίντεο αποτελούνται από bitmapped ψηφιακές εικόνες, και η ροή ήχου είναι επίσης ψηφιακή.

Τα βασικά χαρακτηριστικά ενός βίντεο είναι τα εξής:

- το μέγεθος και το σχήμα μεμονωμένων εικόνων (καρέ)
- η αναλογία στην οποία καταγράφονται οι εικόνες (αναλογία διαστάσεων - aspect ratio), δηλαδή η σχέση μεταξύ του ύψους και του μήκους της εικόνας του καρέ
- οι διαδικασίες συμπίεσης που χρησιμοποιείται για την κωδικοποίηση των πληροφοριών (codecs)
- η ταχύτητα με την οποία οι εικόνες διαδέχονται η μία την άλλη κατά τη διάρκεια της αναπαραγωγής
- η συνολική διάρκεια του προγράμματος
- ο ήχος που συνοδεύει το βίντεο (stereo ή mono)
- το μέγεθος των αρχείων που δημιουργούνται.

Έχουν εμφανιστεί πολλές αναλογίες διαστάσεων του βίντεο (aspect ratio), ωστόσο μπορούμε να τις ταξινομήσουμε σε τρεις βασικές κατηγορίες: (1) 4:3 (πρότυπο τηλεόρασης) (2) 16:9 (widescreen) και (3) 21:9 (inemascope).

4:3 (πρότυπο τηλεόρασης)



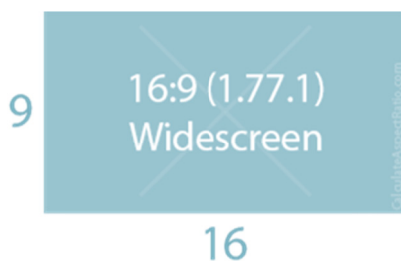
4:3 (1.33:1) πρότυπο τηλεόρασης ήταν σε χρήση από την πρώτη εφεύρεση των καμερών κινούμενης εικόνας. Χρησιμοποιήθηκε για τις ταινίες βωβού κινηματογράφου, οι οποίες αποθηκεύονταν σε φιλμ 35mm. Με την ανακάλυψη της τηλεόρασης, υιοθετήθηκε αυτή η αναλογία για την αποθήκευση μετάδοση εκπομπών και ταινιών στο

τηλεοπτικό κοινό. Είναι επίσης πολύ κοντά στο 1.375:1 αναλογία που είχε οριστεί από την Αμερικανική Ακαδημία Κινηματογραφικών Τεχνών και Επιστημών ως πρότυπο μετά την έλευση του οπτικού ήχου-για-ταινία. Για να διαφοροποιηθεί η βιομηχανία του κινηματογράφου από την τηλεόραση και να προσελκύσει κοινό στους κινηματογράφους, υιοθέτησε την ευρεία αναλογία (widescreen) 16:9. Η αναλογία 16:9, έχει υιοθετηθεί πλέον και στην τηλεόραση και η αναλογία 4:3 πλέον δεν χρησιμοποιείται. Επομένως, τη συναντάμε συχνά σε:

- ✓ 35 mm φιλμ
- ✓ βιντεοκασέτες
- ✓ 1.375:1 φιλμ
- ✓ παλαιές ταινίες και εκπομπές μόνο για τηλεόραση

→ Αναλύσεις: **720x576px, 640x480px, 320x240px**

16:9 (widescreen)



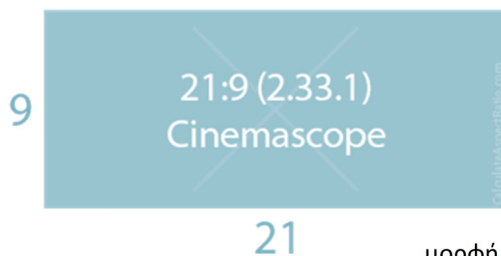
16:9 (1.77:1) ή ευρεία αναλογία (widescreen). Αρχικά, εμφανίστηκε στα κινηματογραφικά φιλμ (1.85:1), για να διαφοροποιήσει τον κινηματογράφο από την τηλεόραση. Πλέον, είναι το πιο διαδεδομένο πρότυπο και χρησιμοποιείται τόσο στο κινηματογράφο όσο και στην τηλεόραση. Είναι το καθιερωμένο πρότυπο για HDTV (High Definition television).

Σχεδόν όλες ψηφιακές κάμερες αποθηκεύουν σε αυτή την αναλογία και αυτή η αναλογία υποστηρίζεται εγγενώς από το πρότυπο DVD. Οι παραγωγοί DVD μπορούν επίσης να επιλέξουν να εμφανίσουν ακόμα πιο ευρείες αναλογίες (cinemascope) όπως 2.39:1. Επόμενος, το συναντάμε συχνά σε:

- ✓ 1.85:1 φιλμ
- ✓ DVD
- ✓ σύγχρονες ταινίες και εκπομπές μόνο για τηλεόραση

→ Αναλύσεις: **1920x1080 px, 1280x720 px, 720x400 px, 640x360 px**

21:9 (cinemascope)



21:9 (2.33:1) Cinemascope. Είναι η κατά προσέγγιση αναλογία διαστάσεων της οθόνης της πραγματικής αξίας 64:27 (2.370:1) σε σύγκριση με την κοινή αναλογία 16:9 (1.77:1). Έχει σχεδιαστεί για ταινίες, οι οποίες καταγράφονται σε σινεμασκόπ ή σύγχρονη αναμορφική μορφή φιλμ 2.39:1.

→ Αναλύσεις: **3840x2160, 1920x1080px, 1680x720px,**

Ανάλογα με το τρόπο που προέκυψε το βίντεο (αναλογικό ή ψηφιακό) και την αναλογία διαστάσεων (aspect ratio), μπορούμε να ορίσουμε τις παρακάτω προδιαγραφές.

Προδιαγραφές για τη διαφύλαξη				
Προέλευση Βίντεο	Αναλογία (aspect ratio)	Προτεινόμενες Αναλύσεις	Μορφότυποι αρχείων	Συμπίεση και κωδικοποίηση (codecs)
Ψηφιακό	16:9	1920x1080 full HD	mp4 (pure) Mpeg	Χρήση συμπίεσης χωρίς απώλειες (lossless): Uncompressed YCbCr
	ή		Mkv	
	21:9			
Ψηφιοποίηση	16:9	720x400	Mp4	JPEG2000 (Lossless)
			mpeg	Huffyuv
Ψηφιοποίηση	4:3	720x576 ή 640x480	Mpeg	Lagarith
			ή avi (για παλαιότερα αρχεία)	
Προδιαγραφές για διάθεση				
Προέλευση Βίντεο	Αναλογία (aspect ratio)	Προτεινόμενες Αναλύσεις	Μορφότυποι αρχείων	Συμπίεση και κωδικοποίηση (codecs)
Ψηφιακό	16:9	1280x720	mp4	Χρήση συμπίεσης με απώλειες (lossless):
	ή			DV
	21:9			
Ψηφιοποίηση	16:9	720x400 ή 640x360	mp4	H.263
				MPEG-2/MPEG-4
Ψηφιοποίηση	4:3	720x576 ή 640x480	Mp4	H.264
				Xvid, DivX

Πίνακας 13: Προδιαγραφές για τα αρχεία βίντεο

Για αρχεία βίντεο, τα οποία είναι ήδη σε ψηφιακή μορφή, συστήνεται η μετατροπή τους σε MPEG4 με κωδικοποίηση H.264 [12]. Αναφορικά με τον ήχο που συνοδεύει το βίντεο, συστήνεται να χρησιμοποιηθούν οι ίδιες προδιαγραφές για τον ήχο, τις οποίες παρουσιάζουμε στην επόμενη ενότητα.

Για οποιονδήποτε κι από τους δύο τρόπος που μπορεί να έχει δημιουργηθεί το αρχείο βίντεο, συστήνεται η χρήση τεχνικών μεταδεδομένων για να εξασφαλιστεί η μακροχρόνια διαφύλαξη του αρχείου [20][23]. Για το σκοπό αυτό μπορεί να χρησιμοποιηθεί είτε το σχήμα μεταδεδομένων PREMIS [21] ή METS [22].

</> Τεχνικά μεταδεδομένα

- ➔ μορφή αρχείου, συμπεριλαμβανομένου εάν η κωδικοποίηση ήταν με ή χωρίς απώλειες (lossy or loseless)
- ➔ πληροφορίες δειγματοληψίας
- ➔ ανάλυση
- ➔ το λογισμικό που χρησιμοποιήθηκε (όνομα και έκδοση)
- ➔ το όνομα της συσκευής που χρησιμοποιήθηκε για την καταγραφή (μάρκα, όνομα, έκδοση μοντέλου)
- ➔ φυσική μορφή του μέσου (είδος και μορφή δίσκου, είδος και μορφή βιντεοκασέτας κλπ)
- ➔ διαστάσεις εξαρτώμενων με το μέσω υλικών (π.χ., διάμετρος, εύρος, ύψος, πάχος, πλάτος, μήκος)
- ➔ πληροφορίες σχετικά με την χρονολογική περίοδο που κατασκευάστηκε το αναλογικό μέσο (τί γενιάς είναι)
- ➔ περιγραφή του πλαισίου (π.χ., αριθμός pixel ή γραμμών μετράει οριζόντια και κάθετα, frame rate)
- ➔ ρυθμός δεδομένων
- ➔ πληροφορίες όσον αφορά τα πρότυπα εκπομπής (π.χ., NTSC, PAL, SECAM)



Διαβάστε επίσης...

- ➔ Οδηγίες ψηφιοποίησης φιλμ από το Federal Agencies Digitization Guidelines Initiative – FADGI (April 18, 2016) http://www.digitizationguidelines.gov/guidelines/Motion_pic_film_scan.html
- ➔ Συγκριτική μελέτη μορφότυπων βίντεο (AVI, MOV, Matroska, MXF, MPEG-2) που χρησιμοποιούνται στην ψηφιοποίηση από το Federal Agencies Digitization Guidelines Initiative – FADGI (December 2, 2014)
http://www.digitizationguidelines.gov/guidelines/FADGI_VideoReFormatCompare_pt1_20141202.pdf
- ➔ Οδηγίες δημιουργίας εξ' αρχής ψηφιακού βίντεο από το Federal Agencies Digitization Guidelines Initiative – FADGI (April 18, 2016)
http://www.digitizationguidelines.gov/guidelines/video_bornDigital.html



2.5 Προδιαγραφές για αρχεία ήχου

Πρόκειται για καταγραφή κατά την οποία ήχοι έχουν αποτυπωθεί με μηχανικό, ηλεκτρομαγνητικό ή ψηφιακό τρόπο σε μέσο που επιτρέπει την αναπαραγωγή του ήχου με τη χρήση ακουστικού εξοπλισμού ή λογισμικού, ώστε να μπορεί ο ήχος να αναπαραχθεί.

Το υλικό αυτό μπορεί να προκύψει από δύο διαφορετικούς τρόπους:

1) Ψηφιοποίηση ήχου από αναλογικά μέσα (κασέτες, δίσκοι βινυλίου κλπ): Πρόκειται για παλιότερο υλικό το οποίο έχει καταγραφεί σε αναλογικά μέσα, όπως μπομπίνες, κασέτες ήχου, δίσκοι βινυλίου κλπ.

2) Ψηφιακοί ήχοι (digital born): Πρόκειται για ήχους οι οποίοι έχουν καταγραφεί χρησιμοποιώντας ψηφιακά μέσα, όπως ψηφιακές κάμερες ή έχουν παραχθεί εξ' αρχής με τη χρήση εξειδικευμένων προγραμμάτων υπολογιστή, π.χ. Sony Creative Software Acid, Apple GarageBand

Και στις δυο περιπτώσεις, η διαδικασία θα πρέπει να οδηγήσει στην παραγωγή τριών τύπων αρχείων:

Αρχείο ήχου για διαφύλαξη	Αρχείο ήχου για διάθεση στο διαδίκτυο	Εικόνα προεπισκόπησης (thumbnail)
1. Ο ήχος στην καλύτερη δυνατή ποιότητα 2. Δεν έχει υποβληθεί σε συμπίεση ή η συμπίεση έχει γίνει χωρίς απώλειες 3. Δεν έχει υποβληθεί σε τροποποιήσεις 4. Αποθηκεύεται κυρίως σε μορφή αρχείου wav ή aiff ή flac 5. Χρησιμοποιείται ο βέλτιστος ρυθμός	1. Χρησιμοποιείται στη θέση του πρωτότυπου για διάθεση στο διαδίκτυο 2. Λογικό μέγεθος αρχείου για γρήγορο χρόνο λήψης, χωρίς να απαιτείται γρήγορη σύνδεση στο διαδίκτυο 3. Συνήθως είναι συμπιεσμένης μορφής 4. Συνήθως αποθηκεύονται	1. Συστήνεται να παράγεται μια εικόνα προεπισκόπησης, η οποία να αντιπροσωπεύει το ήχο ώστε να επιτρέπει στον χρήστη να αποφασίσει αν θέλει να το ακούσει. Αν δεν υπάρχουν θέματα με πνευματικά δικαιώματα, μπορεί να είναι και η εικόνα του cd. 2. Συνήθως αποθηκεύονται σε μορφή GIF ή JPEG μορφές

δειγματοληψίας (συστήνεται 96kHz) και 24 bit βάθος	σε mp3	
	5. Θα πρέπει να έχει 192kbps ή 320 kbps (συστήνεται) ρυθμό μετάδοσης	

Πίνακας 14: Τύποι αρχείων που προκύπτουν από τον ήχο

Τα βασικά χαρακτηριστικά ενός αρχείου ήχου είναι τα εξής:

- **Συχνότητα ή ρυθμός δειγματοληψίας / sample rate (σε kHz):** δείχνει πόσα δείγματα έχουν ληφθεί από τον δειγματολήπτη σε διάρκεια ενός δευτερολέπτου. Πρωταρχικός σκοπός για την επιλογή του ρυθμού δειγματοληψίας είναι η ανάγκη να αποδοθεί η συχνότητα του ακουστικού φάσματος του ανθρώπου, το οποίο είναι περίπου 20kHz. Σύμφωνα με το θεώρημα δειγματοληψίας του Nyquist-Shannon [16][17], για να επιτευχθεί αυτό, ο ρυθμός δειγματοληψίας πρέπει να είναι ο διπλάσιος σε σχέση με την συχνότητα που θέλουμε να επιτύχουμε, δηλαδή τουλάχιστον 40 kHz.
- **Βάθος bit / bit depth:** είναι ο αριθμός των bits που χρησιμοποιούμε για να αποθηκεύουμε το κάθε δείγμα και αντιστοιχεί άμεσα με την ανάλυση κάθε δείγματος. Παραδείγματα βάθος bit περιλαμβάνουν Compact Disc Digital Audio, το οποίο χρησιμοποιεί 16 bits ανά δείγμα, και DVD-Audio και Blu-ray Disc που μπορεί να υποστηρίξει έως και 24 bits ανά δείγμα.
- **Ρυθμός μετάδοσης / Bit rate (σε kbps):** δείχνει τον αριθμό των δυαδικών ψηφίων (bits) που περνάνε από ένα σημείο του καναλιού δεδομένων σε ένα χρονικό διάστημα, όταν αναπαράγεται ο ήχος. Για το αρχείο ήχου, περιγράφει το ποσοστό συμπίεσης του. Όσο πιο μεγάλος είναι ο ρυθμός μετάδοσης, τόσο καλύτερη ποιότητα ήχου έχουμε.

Υπάρχουν πολλά πρότυπα που σχετίζονται με ψηφιακό ήχο. Για τα εμπορικά CD ήχου, οι κατασκευαστές χρησιμοποιούν ρυθμό δειγματοληψίας 44,1 kHz με 16bit βάθος, με σκοπό να αναπαράγουν συχνότητα 22kHz, η οποία είναι κοντά στο όριο των 20kHz, την συχνότητα που ακούει ο άνθρωπος. Ωστόσο τα ψηφιακά αρχεία ήχων που χρησιμοποιούνται σε βιβλιοθήκες και πολιτιστικά ιδρύματα, συστήνεται να χρησιμοποιείται υψηλότερος ρυθμός δειγματοληψίας και βάθος bit, έτσι ώστε να διασφαλίζεται η διαφύλαξη και υψηλή ποιότητα τους, ειδικότερα για ψηφιοποιημένο ήχο από αναλογικό μέσο [17]. Στον παρακάτω πίνακα, προτείνονται οι τύποι αρχείου, ο ρυθμός δειγματοληψίας και το βάθος bit για τα αρχεία διαφύλαξης ήχου [17] [18].

Προδιαγραφές για διαφύλαξη ήχου

Βαθμός Απαίτησης	Μορφότυπος αρχείου	Ρυθμός δειγματοληψίας / sample rate	Βάθος bit / bit depth	Περιγραφή χρήσης
Ελάχιστη	WAV / AIFF / FLAC	44 kHz	16	Ήχος από εμπορικό CD
Μέτρια	WAV / AIFF / FLAC	48 kHz	24	Αφήγηση από άνθρωπο, χωρίς μουσική
Συστήνεται	WAV / AIFF / FLAC	96 kHz	24	Φωνητικές μαρτυρίες, ιδιαίτερα αν υπάρχει και μουσική Φυσικοί ήχοι Ήχοι της φύσης

Προδιαγραφές για διάθεση ήχου

Βαθμός Απαίτησης	Μορφότυπος αρχείου	Ρυθμός δειγματοληψίας / sample rate	Περιγραφή χρήσης
Ελάχιστη	mp3	128 Kbps	Αφήγηση από άνθρωπο, χωρίς μουσική
Μέτρια	mp3	192 Kbps	Μουσική
Συστήνεται	mp3	320 Kbps	Απαίτηση υψηλής πιστότητας

Πίνακας 15: Προδιαγραφές για τα αρχεία ήχου

Αρχεία ήχου τα οποία δεν είναι wav, flac, aiff ή mp3, συστήνεται να μετατρέπονται σε wav format [11][12]:

Αρχικός μορφότυπος (παραδείγματα)	Προτεινόμενος μορφότυπος για ασφαλή διαφύλαξη
<u>WMA</u> : Windows Media Audio File	WAV
<u>RA</u> : Real Audio File	
<u>SND</u> : Apple Sound File	
<u>AU</u> : Sun Audio File	

Πίνακας 16: Προτεινόμενες μετατροπές σε αποδεκτούς μορφότυπους για αρχεία ήχου

Για οποιονδήποτε κι από τους δύο τρόπους που μπορεί να έχει δημιουργηθεί το αρχείο ήχου, συστήνεται η χρήση τεχνικών μεταδεδομένων για να εξασφαλιστεί η μακροχρόνια διαφύλαξη του αρχείου [20] [23]. Για το σκοπό αυτό μπορεί να χρησιμοποιηθεί είτε το σχήμα μεταδεδομένων PREMIS [21] ή METS [22].

</> Τεχνικά μεταδεδομένα

- ➔ μορφή αρχείου, συμπεριλαμβανομένου εάν η κωδικοποίηση ήταν με ή χωρίς απώλειες (lossy or loseless)
- ➔ πληροφορίες δειγματοληψίας
- ➔ αριθμός καναλιών
- ➔ το λογισμικό που χρησιμοποιήθηκε (όνομα και έκδοση)
- ➔ το όνομα της συσκευής που χρησιμοποιήθηκε για την καταγραφή (μάρκα, όνομα, έκδοση μοντέλου)
- ➔ φυσική μορφή του μέσου (είδος και μορφή δίσκου, είδος και μορφή κασέτας κλπ)
- ➔ για αναλογικούς δίσκους, πληροφορίες σχετικά με το τύπο αυλάκωσης (groove)
- ➔ για μαγνητικές ταινίες, πληροφορίες σχετικά με τη τροχιά (track)
- ➔ ταχύτητα και πιθανές προσαρμογές της ταχύτητας για τα αναλογικά μέσα



2.6 Προδιαγραφές για σύνολα δεδομένων

Πρόκειται για υλικό που αποτελείται από δεδομένα και προορίζεται για χρήση ή/και επεξεργασία από υπολογιστή ή και προγράμματα υπολογιστή π.χ. ηλεκτρονικό παιχνίδι, online σύστημα ή υπηρεσία, διαδραστική πηγή, σύνολο δεδομένων, λογισμικό κ.λπ.. Τα σύνολα δεδομένων είναι ήδη σε κάποια δομημένη μορφή (π.χ. πίνακας, xml κλπ.) καθώς πρέπει να δομούνται με τέτοιο τρόπο ώστε να γίνονται κατανοητά από τις υπολογιστικές μηχανές.

Ο τύπος «Σύνολα δεδομένων» περιλαμβάνει επιπλέον τεκμήρια με πρωτογενή ερευνητικά δεδομένα για κάθε έναν από τους άλλους τύπους περιεχομένου π.χ. χαρτογραφικά δεδομένα, εργαστηριακές σημειώσεις, παρουσιάσεις περιπτώσιολογικών μελετών, εννοιολογικά μοντέλα, ροές εργασιών κ.λπ.

Τα σύνολα δεδομένων είναι εξ' αρχής ψηφιακό υλικό. Για να διασφαλιστεί η διαφύλαξη τους, συστήνεται:

- 1) να αποθηκεύονται σε μορφώττυπους ανοικτού τύπου οι οποίοι είναι ανεξάρτητοι λειτουργικού συστήματος και λογισμικού
- 2) να αποθηκεύονται σε δομημένη μορφή (π.χ. xml-based, csv κλπ)
- 3) η πληροφορία να είναι κωδικοποιημένη σε UTF-8.

Οι προδιαγραφές συνοψίζονται στον παρακάτω πίνακα:

Προτεινόμενοι μορφώττυποι για σύνολα δεδομένων	Κωδικοποίηση
<u>CSV</u> : Comma Separated Values File	UTF-8
<u>SIARD</u> : Software Independent Archiving of Relational Databases (open XML format)	
<u>MySQL SQL</u> : Structured Query Language file (MySQL is an open source relational database management system)	
<u>JSON</u> : Javascript Object Notation)	
<u>XLSX</u> : MS Excel Open XML Document (created in MS Office 2007 and 2010)	

Πίνακας 17: Προδιαγραφές για τα σύνολα δεδομένων

Για το αρχείο δεδομένων, συστήνεται η χρήση τεχνικών μεταδεδομένων για να εξασφαλιστεί η μακροχρόνια διαφύλαξη του αρχείου. Για το σκοπό αυτό μπορεί να χρησιμοποιηθεί είτε το σχήμα μεταδεδομένων PREMIS [21] ή METS [22]

</> Τεχνικά μεταδεδομένα

- ➔ μορφή αρχείου
- ➔ κωδικοποίηση της πληροφορίας (π.χ. UTF-8, ANSI κλπ)
- ➔ το λογισμικό που χρησιμοποιήθηκε (όνομα και έκδοση)
- ➔ πιθανές οδηγίες για την προβολή και χρήση των αρχείων

3 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 στη διεύθυνση <http://digitization.hpclab.ceid.upatras.gr> , http://www.infosoc.gr/infosoc/el-GR/services/elibrary/reports_list/psifiopiisi/synoptikos_odigos.htm
- [2] Cornell University Library, "Moving Theory into Practice: Digital Imaging Tutorial", διαθέσιμο στο <http://www.library.cornell.edu/preservation/tutorial/contents.html>
- [3] Harvard University Library, "Selection For Digitizing: A Decision-Making Matrix", <http://www.clir.org/pubs/reports/hazen/matrix.html>
- [4] International Federation of Library Associations and Institutions (IFLA), "Guidelines for digitization projects", διαθέσιμο στο <http://www.ifla.org/VII/s19/pubs/digit-guide.pdf>
- [5] 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
- [6] Ouksel A.M. and Sheth A. (1999) Semantic Interoperability in Global Information Systems, ACM SIGMOD Record, Vol 28(1) March 1999, pp 5-12.
- [7] Sustainability of Digital Formats: Planning for Library of Congress Collections, Content Categories (accessed April 15, 2016) <http://www.digitalpreservation.gov/formats>
- [8] Routhier Perry, S. (2014). Digitization and digital preservation: A review of the literature. SLIS Student Research Journal, 4(1). Retrieved from: <http://scholarworks.sjsu.edu/slissrj/vol4/iss1/4>
- [9] Consortium of Academic and Research Libraries of Illinois (CARLI) (2014), Guidelines for the Creation of Digital Collections: Digitization Best Practices for Two-Dimensional Objects (accessed April 15, 2016) https://www.carli.illinois.edu/sites/files/digital_collections/documentation/guidelines_for_images.pdf
- [10] Consortium of Academic and Research Libraries of Illinois (CARLI) (2014), Guidelines for the Creation of Digital Collections: Digitization Best Practices for Three-Dimensional Objects (accessed April 15, 2016) https://www.carli.illinois.edu/sites/files/digital_collections/documentation/guidelines_for_3D.pdf
- [11] Sustainability of Digital Formats: Planning for Library of Congress Collections, Still Image Preferences in Summary (accessed April 20, 2016) <http://www.digitalpreservation.gov/formats/Still Image >> Preferences in Summary>
- [12] Division of Digital Curation, Shallcross Michael, Deromedi Nancy (2013), Division Records - Digital Curation - Strategies and Procedures - Format Conversion for Long-Term Preservation at the Bentley Historical Library (accessed April 20, 2016) <http://hdl.handle.net/2027.42/96438>
- [13] The Western States Digital Standards Group, Digital Imaging Working Group, Digital Imaging Best Practices, available on the Lyrasis site (<http://www.lyrasis.org>), January 2003
- [14] Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files http://www.digitizationguidelines.gov/guidelines/FADGI_Still_Image-Tech_Guidelines_2010-08-24.pdf , August 2010
- [15] CDL Guidelines for Digital Images (January 2011) http://www.cdlib.org/services/access_publishing/dsc/contribute/docs/cdl_gdi_v2.pdf
- [16] The [Nyquist-Shannon sampling theorem](http://en.wikipedia.org/wiki/Nyquist-Shannon_sampling_theorem) https://en.wikipedia.org/wiki/Nyquist%E2%80%93Shannon_sampling_theorem

- [17] Consortium of Academic and Research Libraries of Illinois (CARLI) (2014), Guidelines for the Creation of Digital Collections: Digitization Best Practices for Audio (accessed April 25, 2016) https://www.carli.illinois.edu/sites/files/digital_collections/documentation/guidelines_for_audio.pdf
- [18] CDP Digital Audio Working Group, Digital Audio Best Practices Version 2.1 <http://sustainableheritagenetwork.org/creator/cdp-digital-audio-working-group>
- [19] Γ. Αγγελάκη, Ανοικτά και Διασυνδεδεμένα Πολιτιστικά Δεδομένα, Αθήνα: Εθνικό Κέντρο Τεκμηρίωσης, 2014 <http://hdl.handle.net/10442/14420>
- [20] Best Practices for Technical Metadata: http://www.library.illinois.edu/dcc/bestpractices/chapter_10_technicalmetadata.html
- [21] PREMIS Metadata: <http://www.loc.gov/standards/premis>
- [22] METS Metadata: <http://www.loc.gov/standards/mets>
- [23] AudioMD and videoMD – Technical Metadata for Audio and Video <https://www.loc.gov/standards/amdvmd/>
- [24] Παναγιώτης Σταθόπουλος, Νίκος Χούσος, και Γεώργιος Σταύρου, “Προδιαγραφές και χαρακτηριστικά διαλειτουργικότητας για ανοικτό ψηφιακό περιεχόμενο.” Μαρ-2013. <http://hdl.handle.net/10442/8887>

4 Κατάλογος Πινάκων και Εικόνων

4.1 Κατάλογος Πινάκων

Πίνακας 1: Τύποι αρχείων εικόνας	12
Πίνακας 2. Ελάχιστες απαιτήσεις για ψηφιοποίηση δισδιάστατων αντικειμένων σύμφωνα με [1]	13
Πίνακας 3: Προδιαγραφές για την Εικόνα Διάθεσης και την Εικόνα Προεπισκόπησης δισδιάστατων αντικειμένων (2d) [9]	14
Πίνακας 4: Προδιαγραφές ψηφιοποίησης με φωτογράφιση [29]	15
Πίνακας 5: Προδιαγραφές για την Εικόνα Διάθεσης και την Εικόνα Προεπισκόπησης τρισδιάστατων αντικειμένων (3d)	15
Πίνακας 6: Προτεινόμενοι μορφότυποι και μετατροπές για εξ' αρχής ψηφιακά αρχεία εικόνας	17
Πίνακας 7: Προδιαγραφές για την Εικόνα Διάθεσης και την Εικόνα Προεπισκόπησης εξ' αρχής ψηφιακών εικόνων	17
Πίνακας 8: Τύποι αρχείων κειμένου	19
Πίνακας 9. Επιπρόσθετες απαιτούμενες προδιαγραφές ψηφιοποίησης	21
Πίνακας 10: Προτεινόμενοι μορφότυποι για εξ' αρχής ψηφιακά αρχεία κειμένου	21
Πίνακας 11: Προτεινόμενες μετατροπές σε αποδεκτούς μορφότυπους για εξ' αρχής ψηφιακά αρχεία κειμένου	22
Πίνακας 12: Τύποι αρχείων που προκύπτουν από το βίντεο	24
Πίνακας 13: Προδιαγραφές για τα αρχεία βίντεο	26
Πίνακας 14: Τύποι αρχείων που προκύπτουν από τον ήχο	30
Πίνακας 15: Προδιαγραφές για τα αρχεία ήχου	31
Πίνακας 16: Προτεινόμενες μετατροπές σε αποδεκτούς μορφότυπους για αρχεία ήχου ..	32
Πίνακας 17: Προδιαγραφές για τα σύνολα δεδομένων	33

4.2 Κατάλογος Εικόνων

Εικόνα 1: Το ταξίδι της πληροφορίας: Από τη δημιουργία του ψηφιακού αρχείου στη διάθεση του στους χρήστες	4
Εικόνα 2: Δομή Οδηγού	9
Εικόνα 3: Τύποι περιεχομένου ΕΚΤ	10



πιστοποίηση προδιαγραφών
διαλειτουργικότητας & περιεχομένου



ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ
www.ekt.gr