

## **Ψηφιοποίηση Υλικού και Δημιουργία Ψηφιακών Βάσεων Δεδομένων**

**Ανέστης Σίτας**

*Δημοσιευμένο στο περιοδικό «Σύγχρονη Βιβλιοθήκη & Υπηρεσίες Πληροφόρησης», τευχ. 14 (Μάιος-Ιούνιος 2002), σ. 19-26*

### **Περίληψη**

Οι ηλεκτρονικές πηγές και οι ψηφιακές βάσεις γίνονται αναπόσπαστο μέρος της συλλογής μιας βιβλιοθήκης. Οι βιβλιοθήκες που διαθέτουν συλλογές που παρουσιάζουν ιδιαίτερο ενδιαφέρον, προγραμματίζουν και σχεδιάζουν την μετατροπή τους σε ψηφιακή μορφή. Στόχος τους, η αξιοποίηση και η ανάδειξη των συλλογών αυτών, με την χρήση των δυνατοτήτων που προσφέρονται από την εφαρμογή των νέων τεχνολογιών. Αυτό το άρθρο περιγράφει την διαδικασία της ψηφιοποίησης ενός αρχείου κειμένου που οδηγεί στην δημιουργία μιας ψηφιακής βάσης με δυνατότητα πρόσβασης μέσα από WWW.

### **Εισαγωγή**

Η ψηφιοποίηση αξιόλογων συλλογών και η διάθεσή τους από το Ιντερνετ, ανοίγει νέους δρόμους περιγραφής, πρόσβασης και ανάκτησης πληροφοριών, που ως τώρα για διαφορετικούς κάθε φορά λόγους, ο εντοπισμός και χρήση τους ήταν από περιορισμένη έως αδύνατη. Τι εννοούμε όμως με τον όρο «ψηφιοποίηση» τεκμηρίων ; Εννοούμε μια σειρά από διαδικασίες που έχουν ως αποτέλεσμα τη μετατροπή αναλογικού υλικού σε ψηφιακή μορφή. Τα τεκμήρια αυτά μπορεί να είναι σχέδια, χάρτες, έργα τέχνης, φωτογραφίες, εφημερα, ή τυπωμένα κείμενα, αλλά ακόμη και άλλα είδη ψηφιακών ή ψηφιοποιημένων τεκμηρίων όπως δεδομένα Η/Υ, βιντεοταινίες και ήχος. Στο άρθρο αυτό θα επικεντρωθούμε στην ψηφιοποίηση αρχείων κειμένου.

### **Αναγκαιότητα**

Η ψηφιοποίηση υλικού στο πλαίσιο μιας ελληνικής βιβλιοθήκης, μπορεί να γίνεται για λόγους ερευνητικούς, πολιτιστικούς ή εκπαιδευτικούς. Έχει ως στόχο την ανάδειξη και διατήρηση - μέσω της ψηφιοποίησης - των συλλογών αυτών και τον

εμπλουτισμό του διαδικτύου με ελληνόγλωσσο υλικό. Η ψηφιοποίηση του υλικού συμβάλει στην ανάδειξή του αφού έτσι γίνεται δυνατή η χρήση του μέσα από ειδικά λογισμικά που δίνουν πολλαπλές δυνατότητες αξιοποίησής του. Ένα ψηφιοποιημένο κείμενο μπορεί να διατεθεί στους χρήστες σε πλήρη μορφή, να γίνει δυνατή η αναζήτησή του με πολλούς τρόπους, να διατεθεί σε διαφορετικές μορφές, αλλά και να γίνει δυνατή η πρόσβασή του εξ αποστάσεως. Μπορεί να χρησιμοποιηθεί για την διαφύλαξή του και την προστασία του από την καταστροφή που μπορεί να προέλθει από την συχνή και ανεξέλεγκτη χρήση του πρωτοτύπου. Σε κάθε περίπτωση πάντως δεν πρέπει να ξεχνάμε ότι η ψηφιοποίηση δεν είναι ούτε διατήρηση ούτε συντήρηση.

#### **Διαχείριση του έργου**

Η διαχείριση του έργου ξεκινά κατά το στάδιο σχεδίασης πριν την έναρξή του. Η σχεδίαση πρέπει να ορίζει τη σκοπιμότητά του και κατά πόσο είναι τεχνικά εφικτό να έλθει εις πέρας (κόστος, προσωπικό, μηχανογραφικός εξοπλισμός και λογισμικά), αλλά και τους ανασταλτικούς παράγοντες που πιθανόν να το εμποδίζουν (έλλειψη δικαιωμάτων χρήσης του υλικού, κτλ). Οι αποφάσεις έργων που βασίζονται σε μεγάλο βαθμό στην τεχνολογία, χρειάζεται να αναθεωρούνται και να επαναπροσδιορίζονται συχνά, καθώς περισσότερες τεχνικές λεπτομέρειες γίνονται διαθέσιμες.

Η καταγραφή και η αξιολόγηση όλων των παραμέτρων που ενέχονται σε ένα τέτοιο σχέδιο, γίνονται με την διεξαγωγή μιας εμπειριστατωμένης και λεπτομερούς οικονομοτεχνικής μελέτης. Όλες οι διαδικασίες που περιγράφονται παρακάτω πρέπει να αναπτυχθούν και να εφαρμοστούν σχολαστικά. Η εφαρμογή τους θα διασφαλίσει την ανάπτυξη ενός ψηφιακού αρχείου υψηλής ποιότητας με εσωτερικούς μηχανισμούς για τη διασφάλιση της μακροχρόνιας επιβίωσης και χρησιμότητάς του.

#### **Πολιτική επιλογής υλικού**

Αν το υλικό προς ψηφιοποίηση δεν έχει μια ομοιόμορφη και ενιαία μορφή ώστε να ψηφιοποιηθεί το σύνολό του ή είναι ένα δυναμικό-ενεργό αρχείο που συνεχώς αλλάζει ή εμπλουτίζεται, είναι αναγκαίο να αποφασιστεί ποιο ακριβώς μέρος του ή μέρη του θα ψηφιοποιηθούν. Για τον λόγο αυτό είναι καλό να δημιουργείται γραπτή πολιτική επιλογής υλικού προς ψηφιοποίηση.

### **Απόκτηση πνευματικών δικαιωμάτων**

Αποτελεί πρωτεύων ζήτημα η εξασφάλιση του δικαιώματος χρήσης του υλικού. Είναι αναγκαία η επικοινωνία με όσους κατέχουν το copyright (ιδιώτες, ιδρύματα, εκδότες) απ' όπου ζητείται η έγγραφη παραχώρηση δικαιωμάτων χρήσης. Σε περίπτωση που το σύνολο ή ένα μέρος του υλικού απαιτεί την απόδοση πνευματικών δικαιωμάτων, έπονται οι συμφωνίες παραχώρησής του. Για τον σκοπό αυτό προχωράμε σε :

- Αναζήτηση όλων των φυσικών ή νομικών προσώπων τα οποία κατέχουν το copyright των εντύπων.
- Οργάνωση αρχείου επιστολών και διαπραγματεύσεων copyright, ώστε να υπάρχουν έντυπα αποδεικτικά στοιχεία παραχώρησής του και οι προσπάθειες εύρεσης της ιδιοκτησίας πνευματικών δικαιωμάτων για παλαιό ή ασαφή υλικό.
- Οι διαπραγματεύσεις για το δικαίωμα χρήσης δεν περιορίζονται μόνο στην δωρεάν παραχώρηση του. Επεκτείνονται και σε άλλους τρόπους χρήσης του, αφού είναι δυνατό να υπάρξουν πολλαπλοί τρόποι διάθεσής του, π.χ. μόνο σε εξουσιοδοτημένους χρήστες, μόνο από συγκεκριμένους υπολογιστές (αναγνώριση ip addresses) ή να υπάρξει χρήση με απόδοση ορισμένου χρηματικού τιμήματος.

Εκτός της αναζήτησης και της παραχώρησης δικαιωμάτων χρήσης στους δικαιούχους των πρωτοτύπων, εγείρονται και ανάλογα ζητήματα που αφορούν το νέο πλέον υλικό που προκύπτει. Μετά την ψηφιοποίηση και την επεξεργασία που έχει υποστεί, υπόκειται με την σειρά του σε ένα σύνολο περιορισμών σχετικά με τα δικαιώματα διάθεσης και χρήσης του. Γι' αυτό πρέπει να φροντίσουμε να διασφαλίσουμε και τα δικαιώματα που αποκτά η βιβλιοθήκη μετά την δημιουργία του νέου υλικού.

### **Επιλογή και προμήθεια λογισμικού**

Υπάρχουν στην αγορά διάφορα προγράμματα διαχείρισης ψηφιακών αρχείων. Εκτός αυτών, κυκλοφορούν στο διαδίκτυο και άλλα που διατίθενται ελεύθερα. Πολλές φορές τα προγράμματα αυτά δεν υπολείπονται σε τίποτε από τα αντίστοιχα εμπορικά. Βέβαια η επιλογή ενός τέτοιου προγράμματος συνεπάγεται και πάλι κάποιο κόστος που θα

προκύψει από την παραμετροποίηση και την εγκατάστασή του. Σε κάθε περίπτωση μια λεπτομερής αναζήτηση όλων των ειδών των διαθέσιμων λογισμικών, του τελικού κόστους τους, αλλά και των τεχνικών δυνατοτήτων τους, πρέπει να συνυπολογίζονται.

#### **Παραμετροποίηση λογισμικού**

Ακολουθεί η προσαρμογή του λογισμικού στις συγκεκριμένες ανάγκες της κάθε βιβλιοθήκης. Είναι σημαντικό να ληφθούν υπόψη εκ των προτέρων οι πιθανές χρήσεις της βάσης δεδομένων προτού καθοριστούν οι ανάγκες αναζήτησης / παρακολούθησης (π.χ. ποια πεδία στη βάση δεδομένων θα ευρετηριαστούν). Ακόμη και οι καλύτερες ψηφιακές συλλογές θα ήταν άχρηστες χωρίς τη δυνατότητα αναζήτησης και ανάκτησης του υλικού τους. Αυτό επιτυγχάνεται με την επιλογή προτύπων λέξεων κλειδιών ή θεματικών επικεφαλίδων αλλά και όλων των άλλων πεδίων που θα καταχωρηθούν στα μεταδεδομένα (metadata). Για τη σωστή λειτουργία του μηχανισμού αναζήτησης μεγάλη σημασία έχει η επάρκεια των μεταδεδομένων που παρέχονται μαζί με το αρχείο, καθώς και η συμβατότητά τους ώστε να είναι δυνατή η συνεργασία μεταξύ διαφόρων ηλεκτρονικών πηγών και η μελλοντική μεταφορά της δημιουργούμενης βάσης σε άλλα πιο εξελιγμένα λογισμικά που θα προκύψουν στο μέλλον. Σημαντικό ρόλο παίζει και η δημιουργία του κατάλληλου interface (αν είναι δυνατόν και ελληνοποίησή του).

#### **Επιλογή χαρακτηριστικών ποιότητας για την ψηφιοποίηση του υλικού**

Ο πιο εύκολος τρόπος ψηφιοποίησης ενός εγγράφου είναι το σκανάρισμα (η σάρωση). Το σκανάρισμα γίνεται με το μηχάνημα οπτικής σάρωσης, το scanner. Παράγοντες που πρέπει να ληφθούν υπόψη στην επιλογή του σαρωτή είναι η οπτική ανάλυση (π.χ. 300, 400 ή 600 dpi = κουκίδες ανά ίντσα), η οπτική πυκνότητα, η περιοχή σάρωσης και ο χρόνος σάρωσης. Κάθε ένας από αυτούς τους παράγοντες επηρεάζει την ποιότητα της τελικής εικόνας. Επίσης τελευταία οι ψηφιακές κάμερες χρησιμοποιούνται ολοένα και περισσότερο για ψηφιοποίηση εικόνας.

Κατά την διάρκεια της αρχικής μελέτης και επιλογής των κατάλληλων λογισμικών και προτύπων, αποφασίζεται σε ποια μορφή θα αποθηκευτεί το ψηφιοποιημένο υλικό, όπως PDF files (χρησιμοποιώντας π.χ. το Adobe Acrobat), σε μορφή GIF ή JPEG, κτλ.

Αλλά και αν προχωρήσουμε σε επεξεργασία με OCR (Optical Character Recognition - Οπτική Αναγνώριση Χαρακτήρων) ή σε έναν συνδυασμό των παραπάνω.

Με το σκανάρισμα δεν μετατρέπεται αυτόματα το κείμενο σε μορφή αναγνώσιμη από τον υπολογιστή. Αυτό που επιτυγχάνεται είναι η δημιουργία μιας ψηφιακής εικόνας (μιας φωτογραφίας) αυτού του εγγράφου. Η μετατροπή της «φωτογραφίας» αυτής σε μορφή που να μπορεί να διαβαστεί από τον υπολογιστή ως κανονικό και πλήρες κείμενο, επιτυγχάνεται με ένα λογισμικό OCR που «διαβάζει» το κείμενο από μια εικόνα ενός τεκμηρίου και την μετατρέπει σε κείμενο ASCII.

#### **Εργασίες ψηφιοποίησης**

Η δημιουργία και παράδοση ψηφιακών αρχείων υλικού μπορεί να διαιρεθεί σε τρεις κύριες περιοχές δραστηριοτήτων:

- Ψηφιοποίηση των δεδομένων (data capture): ψηφιοποιείται το υλικό, προστίθεται κείμενο που τις περιγράφει (metadata) και η συνδυασμένη εγγραφή αποθηκεύεται για μελλοντική επαναχρησιμοποίησή της.
- Πρόσβαση στα δεδομένα και παράδοση: οι διαχειριστές των ψηφιακών αρχείων χρειάζεται να σχεδιάσουν και να εφαρμόσουν έναν μηχανισμό που θα διασφαλίζει πως οι εικόνες θα είναι προσπελάσιμες από τους χρήστες.
- Διαχείριση της συλλογής των δεδομένων: οι δύο προηγούμενες δραστηριότητες πρέπει να επιβλέπονται και να συντονίζονται, ώστε να διασφαλίζεται η εφαρμογή όλων των προτύπων για την παραγωγή μιας ποιοτικής ψηφιακής συλλογής .

#### **Δημιουργία και καταχώρηση μεταδεδομένων**

Η προσθήκη των μεταδεδομένων εκ των υστέρων είναι δύσκολη. Το ψηφιοποιημένο υλικό πρέπει να παρακολουθείται από την πηγή του ως τη ψηφιοποίηση, την προσθήκη μεταδεδομένων, την αρχειοθέτηση, τη διόρθωση και την παράδοση. Ο ποιοτικός έλεγχος του καταχωρημένου υλικού και των μεταδεδομένων του είναι απαραίτητος. Για παράδειγμα, αν το σκαναρισμένο τεκμήριο χρειάζεται διόρθωση, πρέπει να εφαρμοστούν διαδικασίες ελέγχου ώστε να παρέχεται έλεγχος πρόσβασης και έκδοσής του. Ακόμη, ενδεχομένως να απαιτηθεί πιθανή επαναψηφιοποίηση υλικού σε περίπτωση που διαπιστωθούν προβλήματα ευκρίνειας στην εικόνα, κτλ.

### **Αποθήκευση ψηφιοποιημένου υλικού**

Η αποθήκευση του αρχείου περιλαμβάνει τόσο on-line όσο και off-line αποθήκευση. Οι εικόνες υψηλής ποιότητας (masters) είναι προτιμότερο να αποθηκεύονται off-line για λόγους διατήρησης και η πρόσβαση σε αυτές δεν είναι συχνή. Οι εικόνες για πρόσβαση και παράδοση μπορούν να αποθηκεύονται on-line. Αυτές οι «εικόνες πρόσβασης» συνήθως είναι χαμηλότερης ευκρίνειας από τις «εικόνες διατήρησης». Το μέγεθος των εικόνων έχει σημαντικό αντίκτυπο στο μέγεθος του χώρου αποθήκευσης που απαιτείται (και κατά συνέπεια και στο κόστος) και πρέπει να ληφθεί μέριμνα κατά το ξεκίνημα του σχεδίου κατά τον καθορισμό της ανάλυσης όταν θα πραγματοποιείται η σάρωση.

### **Μηχανισμοί αναζήτησης και ανάκτησης**

Η διεπαφή του χρήστη με το σύστημα αναζήτησης πρέπει να σχεδιαστεί με τρόπο που να είναι εύχρηστη και εύκολη. Η βέλτιστη πρακτική είναι η παροχή ενός φάσματος μηχανισμών αναζήτησης για όλους τους χρήστες του συστήματος (προσωπικό και αναγνώστες) ώστε οι αναγνώστες να μην μπορούν να παρεμβαίνουν ή να καταστρέφουν δεδομένα. Ακόμη πρέπει να προσφέρει δυνατότητες αναζήτησης για χρήστες διαφορετικών επιπέδων, για αρχάριους και προχωρημένους.

Το πλεονέκτημα της αποθήκευσης εικόνων σε ψηφιακή μορφή, μετά από την επεξεργασία τους με λογισμικό οπτικής αναγνώρισης χαρακτήρων (OCR), είναι πως καθίσταται δυνατή η πλήρης ανάκτηση του κειμένου και η ευρετηρίαση του συνόλου του. Στην περίπτωση αυτή και εξαιτίας του μεγάλου μεγέθους των ευρετηρίων που δημιουργούνται, η ανάκληση των δεδομένων μπορεί να είναι αργή. Γι' αυτό πιθανώς να είναι προτιμότερος ο περιορισμός της αναζήτησης μόνο σε ορισμένα πεδία. Οι δυνατότητες αναζήτησης για ανάκληση πλήρους κειμένου περιλαμβάνουν τελεστές Bool, stemming, θησαυρούς και wildcards. Το ψηφιοποιημένο κείμενο όταν ανακαλείται μπορεί να διαβαστεί ως πλήρες κείμενο στην οθόνη, να σωθεί στον σκληρό δίσκο του υπολογιστή ή σε δισκέτα και να εκτυπωθεί.

Έχει καταστεί πλέον δυνατή και η αναζήτηση για πληροφορία σε άλλες βάσεις που να υποστηρίζουν πρωτόκολλα όπως Z39.50 ή το WHOIS++. Τέτοια πρωτόκολλα υλοποιούνται στη μηχανή αναζήτησης. Μαζί με τις τυποποιημένες δομές metadata, όπως

το Dublin Core Element Set (DCES), γίνεται ολοένα και περισσότερο δυνατή η συνεργασία μεταξύ των διαφόρων ηλεκτρονικών πηγών. Αυτές οι αποφάσεις πρέπει να ληφθούν σαν μέρος των χαρακτηριστικών λειτουργίας και σχεδίασης του συστήματος.

#### **Διατήρηση και ασφάλεια ψηφιακών αρχείων**

Η ψηφιακή διατήρηση είναι μια διεργασία κατά την οποία οι αποφάσεις που λαμβάνονται για την δημιουργία ενός αρχείου ψηφιακών εικόνων, του παρέχουν την δυνατότητα να είναι χρησιμοποιήσιμο ανεξάρτητα από τις μελλοντικές αλλαγές στην τεχνολογία υλικού και λογισμικού. Η ασφάλεια αφορά στο να καταστεί το αρχείο ψηφιακών εικόνων ασφαλές από πληθώρα απειλών και προβλημάτων. Πάντως, είναι απαραίτητη μια ισορροπία μεταξύ των απαιτήσεων ασφάλειας και των αναγκών των χρηστών. Για παράδειγμα, εκτεταμένη ασφάλεια θα μπορούσε να αποθαρρύνει πιθανούς χρήστες που δεν έχουν μεγάλη εμπειρία στη χρήση δικτυακών πηγών.

#### **Εγγραφή του χρήστη και παροχή αδειών**

Η εγγραφή του χρήστη μπορεί να χρησιμοποιηθεί για πληθώρα σκοπών όπως είναι οι πρακτικές υδατογράφησης (watermarking) και η αυθεντικοποίηση, που προσθέτουν μια επιπλέον διάσταση ασφάλειας στην βάση. Μπορεί να χρησιμοποιηθεί για την διασφάλιση πως ο χρήστης έχει διαβάσει και συμφωνήσει με την παροχή της άδειας και το copyright του υλικού ώστε να δεσμεύεται σε αυτά. Η παροχή αδειών χρήσης του ψηφιοποιημένου υλικού είναι ένας ακόμη παράγοντας που πρέπει να ληφθεί υπόψη, ιδιαίτερα όταν αυτό βρίσκεται υπό καθεστώς copyright. Το να έχει ο χρήστης συνείδηση της συμφωνίας για άδεια πρόσβασης είναι ύψιστης σημασίας και αυτό μπορεί να επιτευχθεί με την τοποθέτηση μιας ειδοποίησης σχετικά με το copyright της συλλογής στην ιστοσελίδα διάθεσής του που να περιέχει ακριβείς πληροφορίες σχετικά με το πως και για ποιο λόγο μπορεί να χρησιμοποιηθεί το υλικό αυτό.

#### **Ψηφιακή υδατογράφηση (watermarking)**

Η ευκολία αντιγραφής εικόνων σε ένα δικτυακό περιβάλλον είναι μεγάλη. Μια πιθανή λύση σε αυτό το πρόβλημα είναι η ψηφιακή υδατογράφηση η οποία είναι αποδεκτή σε περιπτώσεις καταπάτησης του copyright. Ένα οπτικό υδατογράφημα μπορεί

απλά να περιλαμβάνει την τοποθέτηση ενός λογοτύπου, με τρόπο ο οποίος να μην αλλοιώνει την ποιότητα της εικόνας, σε μία γωνία της εικόνας που να δηλώνει την προέλευσή της. Μπορεί να τοποθετηθεί σε κάθε τεκμήριο της βάσης ώστε να ανιχνεύεται αργότερα χωρίς λάθη. Ένας άλλος τρόπος είναι η καταγραφή της ειδοποίησης δικαιώματος χρήσης και άλλων συναφών πληροφοριών σε μερικές επιπλέον γραμμές όταν «κατεβαίνει η εικόνα». Οι χρήστες μπορούν να αφαιρέσουν αυτό το λογότυπο ή την ειδοποίηση δικαιώματος χρήσης, αλλά τότε καταπατούν τη συμφωνία του copyright.

#### **Αυθεντικοποίηση και εξουσιοδότηση**

Όταν έχει αποφασιστεί ότι ένα ορισμένο μέρος του υλικού είναι διαθέσιμο μόνο σε χρήστες που κατέχουν ειδική άδεια, τότε η δυνατότητα να καταλαβαίνει ένας εξυπηρετητής (διακομιστής), πως ο χρήστης είναι αυτός που δηλώνει πως είναι, ονομάζεται αυθεντικοποίηση. Έτσι μπορεί να περιορίσει την πρόσβαση σε ορισμένους χρήστες και να επιτρέπει την «είσοδο» μόνο σε όσους έχουν δικαίωμα χρήσης.

Το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης με βάση όλα όσα περιγράφονται εδώ, στο πλαίσιο του Β' ΕΠΕΑΕΚ του Γ' ΚΠΣ, θα προσπαθήσει να δημιουργήσει μία ή αν είναι δυνατόν και δύο τέτοιες βάσεις, που θα περιέχουν υλικό που αφορά τις ανθρωπιστικές επιστήμες.

#### **Βιβλιογραφία**

1. Berger, Marilyn. "Digitization for preservation and access: a case study." *Library Hi Tech*, v. 17, no. 2 (1999), 146-151.
2. Bouche, Nicole. *Digitization for scholarly use: the Boswell papers project at the Beinecke Rare Book and Manuscript Library*. Washington, DC: Council on Library and Information Resources, 1999.
3. Ellis, Judith (επιμ.). *Η διαχείριση των αρχείων*. Αθήνα: Ελληνική Αρχαιακή Εταιρεία, Τυπωθήτω, 2000.
4. Hudgins, Jean & Mackin, Lisa A. "New materials, new process: implementing digital imaging projects into existing workflow." *Library collections, acquisitions, and technical services*, v. 24, no. 2 (Sum. 2000), 189-204.



5. Lesk, Michael. *Practical digital libraries: books, bytes, and bucks*. San Francisco, Cal.: Morgan Kaufman Publishers, 1997.
6. Ράπτης, Πασχάλης. “Πλαίσιο περιγραφής πηγών και μεταπληροφορίες στο διαδίκτυο.” Στο *8<sup>ο</sup> Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών*. Ρόδος: Βιβλιοθήκη Πανεπιστημίου Αιγαίου, 1999, σ. 309-313.
7. Royan, Bruce. “Content creation and rights management: experiences of SCRAN (the Scottish Cultural Resources Access Network).” *Program*, v. 34, no. 2 (Apr. 2000), 131-142.
8. <http://meta-e.uibk.ac.at/> [πρόσβαση 9/3/2002] (The Metadata Engine Project, σχέδιο δημιουργίας λογισμικού)
9. <http://www.hcu.ox.ac.uk/jtap/> [πρόσβαση 9/3/2002] (Ψηφιοποίηση χειρογράφων)
10. <http://www.nlg.gr/epet> [πρόσβαση 9/3/2002] (Ψηφιοποίηση Επετηρίδων Φιλοσοφικών Σχολών Αθηνών και Θεσσαλονίκης)
11. <http://www.scran.ac.uk> [πρόσβαση 9/3/2002] (SCRAN, βάση δεδομένων με έντυπο και multimedia υλικό)
12. <http://www.jstor.org/> [πρόσβαση 9/3/2002] (JSTOR, μη κερδοσκοπικός οργανισμός που ασχολείται με τη μετατροπή εντύπων περιοδικών σε ηλεκτρονική μορφή)