

# PANDORA

## ένα αρχείο διατήρησης ψηφιακών τεκμηρίων

ΔΙΟΝΥΣΗΣ ΚΟΚΚΙΝΟΣ, ΜΑΡΙΑ ΓΑΒΡΙΛΗ

*Σύμφωνα με τη μυθολογία «το κουτί της Πανδώρας» ήταν ένα πιθάρι που περιείχε όλα τα δεινά του κόσμου. Δόθηκε στην Πανδώρα ως γαμήλιο δώρο από τους θεούς, η οποία και το άνοιξε παρά τη διαταγή του Δία, με αποτέλεσμα να ξεχυθούν τα δεινά αυτά στον κόσμο<sup>1</sup>. Αντίθετα με την ιστορική αυτή προσέγγιση για το «κουτί της Πανδώρας», το αρχείο PANDORA της Εθνικής Βιβλιοθήκης της Αυστραλίας περιλαμβάνει έναν πλούτο και ένα πλήθος πληροφοριών που μελλοντικά θα αποτελέσουν μια πολύτιμη πηγή έρευνας και ενημέρωσης.*

### Εισαγωγή

Η διατήρηση συνιστά ένα πολύ δύσκολο εγχείρημα τόσο στο συμβατικό όσο και στον ψηφιακό κόσμο. Η διατήρηση της εθνικής κληρονομιάς και του πλούτου κάθε χώρας ανήκει συνήθως στις δραστηριότητες των εθνικών βιβλιοθηκών. Αποστολή των εθνικών βιβλιοθηκών είναι η διάσωση των εκδόσεων και των δημοσιευμάτων σε συμβατική αλλά και σε ηλεκτρονική μορφή που αποτελεί ταυτόχρονα και μια μεγάλη πρόκληση. Ήδη από το 1996 η Εθνική Βιβλιοθήκη της Αυστραλίας (National Library of Australia-NLA) προχώρησε στην ανάπτυξη του προγράμματος PANDORA με σκοπό τη διατήρηση των σημαντικών ηλεκτρονικών δημοσιευμάτων που αφορούν στην Αυστραλία και τη διασφάλιση της μακροχρόνιας πρόσβασης σε αυτά.

Το PANDORA αποτελεί ίσως την πιο ολοκληρωμένη προσπάθεια για τη διατήρηση των δημοσιευμάτων σε παγκόσμια κλίμακα. Βασίστηκε στην επιλεκτική προσέγγιση την οποία και υλοποίησε με

τη βοήθεια κατάλληλης επιτροπής, της Selection Committee On-line Australian Publications (SCOAP). Στην προσέγγιση αυτή χρησιμοποιούνται πολιτικές που καθορίζουν τα κριτήρια επιλογής και κατά συνέπεια, τους δικτυακούς τόπους που θα αρχειοθετηθούν. Η SCOAP συστάθηκε πριν την έναρξη της λειτουργίας του PANDORA με σκοπό να αναπτύξει τις κατευθυντήριες γραμμές για την επιλογή των τίτλων που θα εντάσσονταν σε διαδικασίες διατήρησης. Παράλληλα η Εθνική Βιβλιοθήκη της Αυστραλίας προχώρησε στην ίδρυση της Ηλεκτρονικής Μονάδας (Electronic Unit) για την υλοποίηση των προτάσεων της επιτροπής SCOAP. Οι ενέργειες αυτές έθεσαν τις βάσεις για την ανάπτυξη ενός βιώσιμου και τόσο σημαντικού προγράμματος διατήρησης, του PANDORA.

### Γενικά

Το PANDORA (Preserving and Accessing Networked Documentary Resources of Australia)<sup>2</sup> έχει αναπτυχθεί από την Εθνική Βιβλιοθήκη της Αυστρα-

1. Μπαμπινιώτης, Γ., 1998. Λήμμα «Πανδώρα», *Λεξικό της Νέας Ελληνικής Γλώσσας*, σσ. 1322.

2. PANDORA. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/> (Ημερομηνία πρόσβασης: 19.01.05)

λίας (NLA) σε συνεργασία με μια σειρά από άλλους φορείς. Ξεκίνησε το 1996, όταν πλέον είχε γίνει ξεκάθαρο ότι μεγάλος αριθμός αυστραλιανών πληροφοριών ήταν διαθέσιμος μέσω Web. Η Εθνική Βιβλιοθήκη της Αυστραλίας ανέπτυξε την πολιτική, τις κατάλληλες διαδικασίες και ένα σύστημα ψηφιακής αρχειοθέτησης για τη δημιουργία και διαχείριση ενός αρχείου με Web εκδόσεις, όταν η μόνη αντίστοιχη προσπάθεια διεθνώς, ήταν από την Εθνική Βιβλιοθήκη του Καναδά.

Από την αρχή της δημιουργίας του PANDORA, η NLA δίνοντας έμφαση στο συνεργατικό μοντέλο προσέγγισης μιας και το υλικό προς διατήρηση ήταν πολύ μεγάλο σε όγκο για να συγκεντρωθεί μόνο από έναν φορέα, προσκάλεσε την πολιτεία, τις βιβλιοθήκες και σχετικούς πολιτιστικούς φορείς να συνεργαστούν για τη συγκρότηση του αρχείου. Η πρώτη Βιβλιοθήκη που ανταποκρίθηκε ήταν η Κρατική Βιβλιοθήκη της Βικτορίας μόλις το 1998 και ακολούθησαν μια σειρά από άλλους φορείς. Αντίθετα, η Κρατική Βιβλιοθήκη της Τασμανίας (State Library of Tasmania) ανέπτυξε το δικό της Web αρχείο με την ονομασία "Our Digital Island" και ανέλαβε την ευθύνη για την αρχειοθέτηση όλων των εκδόσεων της Τασμανίας. Ωστόσο, μεταξύ των αρχείων της Κρατικής Βιβλιοθήκης της Τασμανίας και της Εθνικής Βιβλιοθήκης της Αυστραλίας υπάρχουν ισχυροί δεσμοί συνεργασίας.

Ο σκοπός<sup>3</sup> του PANDORA είναι να διατηρήσει μόνο, τα δημοσιεύματα που α-

φορούν την Αυστραλία, γνωρίζοντας ότι είναι ανέφικτο να διατηρηθεί όλος ο όγκος των δημοσιευμάτων. Το έργο έχει ως αντικείμενο την αρχειοθέτηση και διατήρηση για τις μελλοντικές γενιές του υλικού που είναι διαθέσιμο στον Αυστραλιανό Ιστό. Είναι σημαντικό να αναφερθεί ότι μέρος του υλικού που έχει αρχειοθετηθεί και είναι διαθέσιμο πλέον από το PANDORA, έχει ήδη εξαφανιστεί από τον Ιστό. Η τάση και ο σκοπός του PANDORA είναι σε κάθε περίπτωση η διατήρηση του σημαντικού περιεχομένου που έχει ή θα έχει αξία για τη χώρα της Αυστραλίας. Δεν αποβλέπει στη διατήρηση της τεχνολογίας ή στον πειραματισμό σε διάφορες τεχνολογίες αλλά στη διατήρηση του περιεχομένου και στη διασφάλιση της ελεύθερης πρόσβασης σε αυτό.

### PANDORA - Selective Archive

Το PANDORA είναι ένα επιλεκτικό αρχείο<sup>4</sup> (selective archive), το οποίο είναι ελεύθερα προσβάσιμο μέσω του Web. Η Εθνική Βιβλιοθήκη της Αυστραλίας και κάθε συνεργαζόμενος φορέας επιλέγει τα δημοσιεύματα που θα αρχειοθετήσει με βάση συγκεκριμένες οδηγίες. Σημειώνεται ότι περιέχονται και εμπορικές εκδόσεις με τη διαφορά ότι υπάρχουν περιόδοι περιορισμένης πρόσβασης για την προστασία των εμπορικών συμφερόντων των εκδοτών, κατά τη διάρκεια των οποίων οι τίτλοι είναι διαθέσιμοι και προσβάσιμοι μόνο από έναν υπολογιστή του αναγνωστηρίου του κάθε συνεργαζόμενου φορέα.

3. Hodge, G.M., 2000. Best Practices for Digital Archiving: An Information Life Cycle. In: *D-Lib Magazine*, vol. 6, no. 1. Διαθέσιμο στη διεύθυνση: <http://www.dlib.org/dlib/january00/01hodge.html> (Ημερομηνία πρόσβασης: 19.01.05)

4. NLA, 2003. PANDORA, Australia's Web Archive: an Overview. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au.background.html> (Ημερομηνία πρόσβασης: 19.01.05)

Η μέθοδος επιλογής παρουσιάζει τα εξής πλεονεκτήματα:

- Κάθε αντικείμενο του αρχείου αποτιμάται ποιοτικά και λειτουργικά. Κάθε αντικείμενο τυγχάνει ποιοτικής αντιμετώπισης στο βαθμό που επιτρέπεται και σε σχέση με τις τρέχουσες τεχνικές δυνατότητες.
- Κάθε αντικείμενο του αρχείου μπορεί να καταλογογραφηθεί πλήρως και επομένως να αποτελεί μέρος της εθνικής βιβλιογραφίας.
- Παροχή δημόσιας πρόσβασης σε κάθε αντικείμενο του αρχείου. Όλα τα αντικείμενα του PANDORA είναι προσβάσιμα, εκτός από περιπτώσεις ενός μικρού αριθμού εμπορικών εκδόσεων.
- Ιστοσελίδες που δεν είναι προσβάσιμες από τις μηχανές (harvesting robots) μπορούν να αρχειοθετηθούν με τη συνεργασία του εκδότη. Συμπεριλαμβάνονται και οι εμπορικές σελίδες που προστατεύονται με κωδικούς πρόσβασης καθώς και εκδόσεις που είναι δομημένες ως βάσεις δεδομένων.

Ωστόσο, η Εθνική Βιβλιοθήκη παραδέχεται ότι υπάρχουν και μειονεκτήματα με αυτήν τη μέθοδο που προκύπτουν κυρίως από τη διαδικασία επιλογής. Η μέθοδος επιλογής των δημοσιευμάτων στηρίζεται σε υποκειμενικές κρίσεις για την αξία που θα έχουν στο μέλλον. Είναι αναπόφευκτο σε ένα επιλεκτικό αρχείο να παραλειφθούν σημαντικές εκδόσεις. Η μέθοδος επιλογής απαιτεί εντατική εργασία (labour-intensive) και το κόστος είναι πολύ υψηλό. Επιπρόσθετα, η Εθνική Βιβλιοθήκη της Αυστραλίας παρακολουθεί συνεχώς τις αντίστοιχες διεθνείς προσπάθειες αναφορικά με τις μεθόδους επιλογής, με σκοπό τη βελτίωση της μεθόδου.

## Αντικειμενικοί στόχοι

Οι αντικειμενικοί στόχοι του PANDORA είναι οι ακόλουθοι:

- Αναγνώριση, επιλογή και καταλογογράφηση των σημαντικότερων αυστραλιανών on-line δημοσιευμάτων χρησιμοποιώντας συγκεκριμένες καθιερωμένες οδηγίες.
- Διακανονισμοί με τους εκδότες για την απόκτηση ενός αντιτύπου για λόγους διατήρησης.
- Παροχή διαθέσιμων πληροφοριών στους χρήστες με ευνοϊκούς όρους.
- Μέριμνα για τα εμπορικά συμφέροντα των εκδοτών αναφορικά με την πρόσβαση των απομακρυσμένων χρηστών.
- Διατήρηση του "look-and-feel" των online ηλεκτρονικών εκδόσεων.
- Ενημέρωση των πληροφοριών που σχετίζονται με το αρχείο σε συστηματική βάση και διατήρηση των ημερομηνιών προηγούμενων εκδόσεων.
- Μετατροπή των εκδόσεων του αρχείου σε νέες μορφές όταν τα πρότυπα αλλάζουν.

Επιπρόσθετα, το PANDORA επεκτείνεται και δραστηριοποιείται σε τομείς όπως:

- Συμμετοχή σε συζητήσεις για επέκταση νομικών θεμάτων και πνευματικής ιδιοκτησίας (copyright) των online δημοσιευμάτων.
- Υλοποίηση ενός συστήματος μόνιμων ονομάτων για τις αυστραλιανές εκδόσεις, με σκοπό να ξεπεραστούν τοπικά προβλήματα που σχετίζονται με μη υποστηριζόμενες συνδέσεις ή συνδέσμους (links).
- Υλοποίηση ενός συστήματος περιγραφής αρχειακών εγγράφων βασισμένο στα στοιχεία του Dublin Core, με σκο-

πό τη βελτίωση και την αποτελεσματικότητα της on-line αναζήτησης.

- Συνεργασία και με άλλες βιβλιοθήκες της Αυστραλίας για την καθιέρωση του PANDORA.

### Περιεχόμενο του PANDORA

Το αρχείο PANDORA διαθέτει μια μεγάλη ποικιλία ηλεκτρονικών εκδόσεων μερικά από τα οποία ανταγωνίζονται την έντυπη μορφή. Επιπλέον, το PANDORA έχει διάφορα είδη εκδόσεων που δεν υπάρχουν στην έντυπη μορφή, όπως sites οργανισμών και προσωπικά. Αναλυτικά, τα είδη των δημοσιευμάτων<sup>5</sup> του αρχείου εντάσσονται στις εξής κατηγορίες:

- Μονογραφίες με πάγιο περιεχόμενο όπως οι έντυπες εκδόσεις.
- Συσσωρευμένες (cumulative) ή εξελισσόμενες (evolving) μονογραφίες, το περιεχόμενο των οποίων αλλάζει.
- Περιοδικές εκδόσεις με σταθερή συχνότητα, τα τεύχη των οποίων παραλαμβάνονται διαδοχικά όπως και στην έντυπη μορφή.
- Εξελισσόμενες περιοδικές εκδόσεις (evolving serials), των οποίων τα περιεχόμενα αλλάζουν με τον χρόνο.
- Οποιοδήποτε δημοσίευμα έχει ISBN, ISSN ή ISMN.
- Εφημερίδες (newspapers sampled in snapshots).
- Αποκόμματα εφημερίδων (ephemera).
- Sites οργανισμών & προσωπικά.
- Χάρτες.
- Ετήσιες αναφορές.
- Βάσεις Δεδομένων.

- Ουσιώδη λογοτεχνικά έργα.
- Πρακτικά συνεδρίων.

Αντίθετα, οι ακόλουθες κατηγορίες<sup>6</sup> αποκλείονται από το αρχείο:

- Ημερήσιες ηλεκτρονικές εφημερίδες.
- Promotional sites και διαφημιστικά.
- Λίστες συζητήσεων (discussion lists), δωμάτια συζήτησης (chat rooms), bulletin boards and news groups.
- CAMS.
- Blogs εκτός από εκείνα που υποστηρίζουν τις ακαδημαϊκές εκδόσεις.
- Παιχνίδια.
- Πύλες πρόσβασης (Portals).
- Μεταπτυχιακές εργασίες (Theses).
- Drafts και έργα σε εξέλιξη.
- Μεμονωμένα άρθρα.
- Δημοσιεύματα μη διαθέσιμα στο Internet.

Το περιεχόμενο του PANDORA οργανώνεται κάτω από 15 κύριες κατηγορίες, οι οποίες είναι οι εξής:

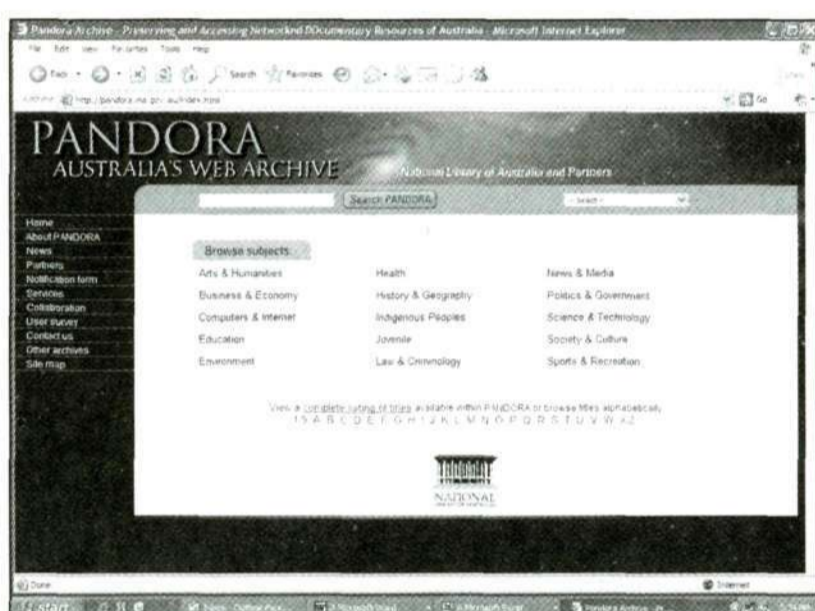
- Τέχνες & Ανθρωπιστικές επιστήμες (Arts & Humanities).
- Επιχειρήσεις & Οικονομία (Business & Economy).
- Υπολογιστές & Διαδίκτυο (Computers & Internet).
- Εκπαίδευση (Education).
- Περιβάλλον (Environment).
- Υγεία (Health).
- Ιστορία & Γεωγραφία (History & Geography).
- Αυτόχθονες άνθρωποι (Indigenous People).
- Νεολαία (Juvenile).

5. NLA, 2003. Business Process Model: 1.3 Online Electronic Publications. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/bpm.html#onli> (Ημερομηνία πρόσβασης: 15.01.05)

6. Phillips, M.E., 2003. Collecting Australian Online Publications-Balanced Scorecard Initiative 49. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/BSC49.doc> (Ημερομηνία πρόσβασης: 16.01.05)

- Δίκαιο & Εγκληματολογία (Law & Criminology).
- Ειδήσεις & ΜΜΕ (News & Media).
- Πολιτική & Κυβέρνηση (Politics & Government).
- Επιστήμη & Τεχνολογία (Science & Technology).
- Κοινωνία & Πολιτισμός (Society & Culture).
- Αθλητισμός & Ψυχαγωγία (Sports & Recreation).

Εσωτερικά σε κάθε κατηγορία το υλικό παρατίθεται αλφαβητικά κατά τίτλο.



Η αρχική σελίδα του PANDORA

## Μέθοδος Επιλογής Ηλεκτρονικών Δημοσιευμάτων

Η Ηλεκτρονική Μονάδα δημιουργήθηκε το 1996 με σκοπό να συμβάλει στην αποτελεσματική ανάπτυξη και υλοποίηση της πολιτικής και των κατάλληλων διαδικασιών που θα εξασφαλίσουν τη μακροχρόνια πρόσβαση στα ηλεκτρονικά δημοσιεύματα. Είναι υπεύθυνη για όλες τις πτυχές που σχετίζονται με τη διαχείριση των ηλεκτρονικών δημοσιευμάτων περιλαμβάνοντας τη μέθοδο επιλογής, τις διαπραγματεύσεις με τους εκδότες, την καταλογογράφηση και την αρχειοθέτηση.

Ειδικότερα, η μέθοδος επιλογής περιλαμβάνει τις ακόλουθες τέσσερις διαδικασίες:

- **Διαδικασία εγγραφής μεταδεδομένων (Metadata Recording Process).** Τα μεταδεδομένα πρέπει να καταχωρούνται στο αρχείο οποιαδήποτε στιγμή κατά τη διάρκεια είτε της αναγνώρισης, επιλογής και αξιολόγησης είτε της αρχειοθέτησης.
- **Διαδικασία αναγνώρισης (Identification Process).** Η Ηλεκτρονική Μονάδα ερευνά τακτικά το διαδίκτυο για αναγνώριση ηλεκτρονικών δημοσιευμάτων με σκοπό τη διατήρησή τους. Επιπρόσθετα, προτάσεις προέρχονται και από άλλα μέλη του προσωπικού καθώς και από ερεθίσματα προερχόμενα από την αίτηση των εκδοτών για ISBN, ISSN ή Cataloging in Print (CIP) που προκαλούν την προσοχή της Ηλεκτρονικής Μονάδας. Όταν εντοπιστεί μια δημοσίευση, η πρώτη εργασία είναι ο έλεγχος των λεπτομερειών έκδοσης με σκοπό να καθοριστεί αν το συγκεκριμένο δημοσίευμα είναι κατάλληλο για αρχειοθέτηση. Έτσι, γίνονται οι ερωτήσεις αν το δημοσίευμα έχει ήδη αποτιμηθεί, αν υπάρχει σε έντυπη μορφή ή σε κάποια άλλη μορφή διαθέσιμο στη συλλογή της Βιβλιοθήκης ή αν ο εκδότης παράγει μια ισότιμη έντυπη έκδοση. Η επεξεργασία των απαντήσεων θα καταλήξει στο αν το δημοσίευμα πρέπει να διατηρηθεί ή όχι.
- **Διαδικασία επιλογής (Selection Process).** Η Ηλεκτρονική Μονάδα αποτιμά κάθε δημοσίευμα έναντι των οδηγιών SCOAP<sup>7</sup>. Τα αποτελέσματα αυτής της διαδικασίας είναι είτε το δημοσίευμα να επιλεγεί ή να απορριφθεί είτε το δημοσίευμα να μην περιέχει επαρκείς πληροφορίες για να ληφθεί μια απόφαση. Στην τελευταία περίπτωση το δημοσίευμα κρατείται σε κατάσταση αναμονής.
- **Διαδικασία Τεχνικής Αξιολόγησης (Technical Evaluation Process).** Η δια-

δικασία αυτή αφορά μόνο στα επιλεγμένα δημοσιεύματα για διατήρηση. Κατά τη διάρκεια της τεχνικής αξιολόγησης συντελούνται τα εξής:

- **Καθορισμός μοναδικών προσδιοριστών** (Allocation of Unique Identifiers). Όπου υπάρχει ο μοναδικός προσδιοριστής (ISBN, ISSN, κ.τ.λ.) θα αναγραφεί στον κατάλογο και στο αρχείο διαχείρισης. Επιπροσθέτως, σε κάθε δημοσίευμα προσδιορίζεται ένα PURL (Persistent Universal Resource Locator) για τη σταθερότητα του δημοσιεύματος και ένας εσωτερικός προσδιοριστής για να επιτρέψει στους διαχειριστές του αρχείου την εύκολη ανάκτησή του.
- **Εγγραφή περιγραφικών μεταδεδομένων** (Recording Descriptive Metadata). Τα περιγραφικά μεταδεδομένα που χρησιμοποιούνται για το δημοσίευμα είναι ο τίτλος, ο εκδότης, ο δημιουργός, τα (P)URLs, ο προσδιοριστής, η συχνότητα έκδοσης, αν υπάρχει, διάρκεια ζωής (life date) και άλλες πληροφορίες όπως για παράδειγμα τυχόν περιορισμοί πρόσβασης.
- **Επικοινωνία με τον εκδότη** (Contact with the Publisher). Γίνονται διάφορες διαπραγματεύσεις με τους εκδότες για την άδεια αρχειοθέτησης των δημοσιευμάτων, τους όρους πρόσβασης, κ.ά.
- **Καταλογογράφηση** (Cataloguing). Η Ηλεκτρονική Μονάδα δημιουργεί μια εγγραφή στην εθνική βιβλιογραφία (National Bibliographic Utility Record) περιλαμβάνοντας και τα περιγραφικά μεταδεδομένα. Η κατα-

λογογράφηση βασίζεται στους Anglo-American Cataloguing Rules 2 (AACR2), στις Library of Congress Subject Headings (LCSH) και στα ABN Standards Committee Guidelines.

- **Τεχνική αξιολόγηση** (Technical Evaluation). Στη διαδικασία της τεχνικής αξιολόγησης μπορεί να προστεθούν επιπλέον μεταδεδομένα όπως σύνδεσμοι, το (P)URL του εκδότη, τρέχοντα τεύχη αν πρόκειται για περιοδική έκδοση, κ.ά.

### Οδηγίες επιλογής ηλεκτρονικών δημοσιευμάτων

Η μέθοδος επιλογής των ηλεκτρονικών δημοσιευμάτων παραμένει η πιο βιώσιμη μέχρι και σήμερα, παρέχοντας στην Εθνική Βιβλιοθήκη της Αυστραλίας όλες τις προϋποθέσεις για ένα αρχείο υψηλής ποιότητας με ηλεκτρονικά δημοσιεύματα αναζητήσιμα, προσβάσιμα και λειτουργικά για τους σημερινούς αλλά και τους μελλοντικούς ερευνητές. Κατά το χρονικό διάστημα 2002-2003 επανεκδόθηκαν<sup>8</sup> οι οδηγίες επιλογής του υλικού, οι οποίες επιμελήθηκαν από το Collection Development Management Committee (CDMC). Το CDMC πρότεινε ότι έπρεπε να γίνει μια επιλογή ανάμεσα σε μια συλλογή με μια ευρεία ποικιλία δημοσιευμάτων ή εστίαση σε μια συλλογή σε συγκεκριμένες θεματικές ενότητες με κάλυψη σε βάθος. Κάθε φορέας που συμμετέχει στο PANDORA επιλέγει το υλικό σύμφωνα με τις δικές του οδηγίες επιλογής. Όλα όμως τα δημοσιεύματα αποθη-

7. SCOAP (Selection Committee on Online Australian Publications). Διαθέσιμο στη διεύθυνση: <http://www.nla.gov.au/scoap/> (Ημερομηνία πρόσβασης: 17.01.05)

8. NLA, 2003. Online Australia Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/selectionguidelines.html> (Ημερομηνία πρόσβασης: 17.01.05)

κεύονται κεντρικά σε έναν server της Εθνικής Βιβλιοθήκης.

Κάποιες γενικές οδηγίες επιλογής<sup>9</sup> είναι:

- Όπου υπάρχει έντυπη και online έκδοση, προτιμητέα είναι η έντυπη έκδοση όπου και θα αποκτηθεί για την Εθνική Συλλογή, ενώ η online θα επιλεγεί μόνο για διατήρηση και εφόσον έχει κάποια σημαντική και επιπρόσθετη αξία ή πληροφοριακό περιεχόμενο.
- Όπου υπάρχει η έκδοση σε online μορφή, CD ROM ή δισκέτα προτιμητέα είναι η online μορφή, γιατί οι δυσκολίες διατήρησης είναι λιγότερες.
- Δεν διατηρούνται όλες οι εκδόσεις ενός ηλεκτρονικού δημοσιεύματος.
- Η μορφή ενός ηλεκτρονικού δημοσιεύματος δεν αποτελεί εμπόδιο για την επιλογή του ή όχι. Αυτό σημαίνει ότι υπάρχουν επιλεγμένα δημοσιεύματα που δεν έχουν αρχειοθετηθεί προς το παρόν για τεχνικούς λόγους, όπως τα datasets, ωστόσο η Εθνική Βιβλιοθήκη αναζητεί λύσεις για να ξεπεραστούν τέτοιες δυσκολίες.

Κάποιες ειδικές οδηγίες επιλογής στηρίζονται στο αυστραλιανό περιεχόμενο, δηλαδή ένα σημαντικό τμήμα του δημοσιεύματος πρέπει να αφορά στην Αυστραλία, στις πολλαπλές εκδόσεις και συγκεκριμένα όταν ένα δημοσίευμα είναι διαθέσιμο και στην έντυπη και στην ηλεκτρονική μορφή, η ηλεκτρονική μορφή θα διατηρείται μόνο αν έχει σημαντική επιπρόσθετη αξία και πληροφοριακό περιεχόμενο. Επιπλέον, στο δικαίωμα (authority) και στην ερευνητική αξία του δημοσιεύματος και τέλος στην οριοθέτη-

ση παραμέτρων τίτλων, δηλαδή εάν ένα δημοσίευμα έχει έναν αριθμό εσωτερικών και εξωτερικών συνδέσμων πρέπει να αποφασιστούν τα όρια. Οι εσωτερικοί σύνδεσμοι είναι εκείνοι που θα αρχειοθετηθούν.

## Διαχείριση αρχείου PANDORA

Οι κύριες γενικές δραστηριότητες<sup>10</sup> της διαχείρισης του αρχείου είναι:

- **Πρόσθεση νέων δημοσιευμάτων.** Η διαδικασία αυτή αναφέρεται στη δημιουργία της εγγραφής στο National Bibliographic Utility, στον έλεγχο του PURL για να εξακριβωθεί αν τα νέα δημοσιεύματα έχουν εγγραφεί, τη συλλογή των μεταδεδομένων, τον καθορισμό της κατάλληλης δομής αρχείου και τις εκδόσεις του, την ενεργοποίηση του λογισμικού συγκομιδής (harvesting), τη δημιουργία της σελίδας τίτλου (title entry page) και την ενημέρωση του εκδότη για τη διαθεσιμότητα των νέων συνδέσμων.
- **Εξουσιοδότηση πρόσβασης στα δημοσιεύματα.** Η διαδικασία αυτή αφορά στην πρόσθεση νέων δημοσιευμάτων στα ευρετήρια των μηχανών αναζήτησης, στη σύλληψη (capture) του λογισμικού που θα υποστηρίζει την αναζήτηση των δημοσιευμάτων, στον έλεγχο του λογισμικού, στη μετατροπή των δημοσιευμάτων που έχουν μη αποδεκτά πρότυπα σε αποδεκτά, σε περιπτώσεις που απαιτείται και ρύθμιση του προφίλ πρόσβασης ανά δημοσίευση.
- **Ενημέρωση των δημοσιευμάτων.** Η διαδικασία αναφέρεται στον εντοπι-

9. Ό.π. σημ. 6.

10. NLA, 2003. PANDORA Business Process Model: 4. The archive management component. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/bpm.html> (Ημερομηνία πρόσβασης: 16.01.05)

σμό χαμένων τευχών των περιοδικών δημοσιευμάτων, στην πρόσθεση νέων τόμων και τευχών των ηλεκτρονικών περιοδικών τίτλων, στην πρόσθεση νέων εκδόσεων των μονογραφιών, και στην ενημέρωση των μεταδεδομένων.

- **Μετατροπή των εκδόσεων του αρχείου.** Οι διαδικασίες μετατροπής γίνονται για να επιβεβαιώσουν την ποιότητα και την αριότητα του περιεχομένου του ψηφιακού αντικειμένου. Αυτές οι διαδικασίες περιλαμβάνουν την απενεργοποίηση των λειτουργιών πληρωμής, όπως οι συνδρομές σε ηλεκτρονική μορφή, την απομάκρυνση συνδέσμων, την απομάκρυνση διπλοεγγραφών με τη διαγραφή των URLs και τη διαγραφή περιττών τμημάτων ενός δημοσιεύματος.
- **Διαγραφή των δημοσιευμάτων από το αρχείο.** Η διαγραφή αφορά στην απομάκρυνση ολόκληρου του δημοσιεύματος από το αρχείο. Η διαγραφή γίνεται λόγω των λαθών συγκομιδής (harvesting errors), νομικών θεμάτων και της ανακάλυψης της έντυπης έκδοσης ενός διατηρημένου ηλεκτρονικού αντικειμένου, γιατί η προτίμηση εστιάζεται στην έντυπη έκδοση. Οι διαδικασίες διαγραφής περιλαμβάνουν την απομάκρυνση του περιεχομένου από το αρχείο με τη διαγραφή των URLs, την ενημέρωση των μεταδεδομένων στην υπηρεσία ανταπόδοσης PURL για την απενεργοποίηση των εκδόσεων του PANDORA PURL και την ενημέρωση των μεταδεδομένων στο αρχείο διαχείρισης έτσι ώστε να αντικατοπτρίζεται η κατάσταση του ψηφιακού αντικειμένου.

Ειδικότερα η διαχείριση<sup>11</sup> του αρχείου αφορά και στις δραστηριότητες της υπηρεσίας ανταπόδοσης PURL (PURL resolver service), στα πρότυπα μεταδεδομένων που θα χρησιμοποιηθούν και ειδικότερα στα περιγραφικά μεταδεδομένα. Η περιγραφή των ψηφιακών αντικειμένων, όπου είναι εφικτή πραγματοποιείται με το πρότυπο μεταδεδομένων Dublin Core Metadata Element Set. Το Dublin Core χρησιμοποιείται ακόμα και για τα διοικητικά μεταδεδομένα και για τα μεταδεδομένα διατήρησης.

### Preservation Metadata for Digital Collection

Η Εθνική Βιβλιοθήκη της Αυστραλίας έχει δημιουργήσει μια προμελέτη (exposure draft) για ένα σχήμα μεταδεδομένων με σκοπό τη διατήρηση των ψηφιακών αντικειμένων με την ονομασία Preservation Metadata for Digital Collection<sup>12</sup>. Η ανάπτυξή του προέκυψε λόγω της απουσίας ενός ικανοποιητικού σχήματος διατήρησης Μεταδεδομένων. Τι είναι όμως τα μεταδεδομένα διατήρησης; Τα μεταδεδομένα διατήρησης ορίζονται ως η πληροφορία ή τα δεδομένα που υποστηρίζουν τη διαδικασία της διατήρησης των ψηφιακών τεκμηρίων. Είναι δηλαδή το βασικό συστατικό των συστημάτων αρχειοθέτησης εξυπηρετώντας την εξασφάλιση της μακρόχρονης πρόσβασης στα ψηφιακά τεκμήρια. Οι «στόχοι» ενός σχήματος μεταδεδομένων διατήρησης είναι να προσδώσουν στο ψηφιακό τεκμήριο βιωσιμότητα, απόδοση (renderability) και η ικανότητα κατανόησης (understandability). Ειδικότερα, μέ-

11. NLA, 2003. PANDORA Business Process Model: 5. Issues of Archive Management.

Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/bpm.html> (Ημερομηνία πρόσβασης: 17.01.05)

12. NLA, 1999. Preservation Metadata for Digital Collections. Διαθέσιμο στη διεύθυνση:

<http://www.nla.gov.au/preserve/pmeta.html> (Ημερομηνία πρόσβασης: 17.01.05)



σα σε ένα αρχείο τα μεταδεδομένα διατήρησης «συνοδεύουν» το ψηφιακό τεκμήριο παρέχοντας σχετικές περιγραφικές, δομικές, διοικητικές και πληροφορικές πνευματικής ιδιοκτησίας. Τα μεταδεδομένα αυτά θα «διατηρηθούν»<sup>13</sup> και θα μεταναστεύσουν από μορφή σε μορφή και από πρότυπο σε πρότυπο.

Η προμελέτη του σχήματος επιδιώκει να καλύψει τη διατήρηση των ψηφιακών συλλογών, τη διαχείριση των αρχείων με online ψηφιακές εκδόσεις και τη διαχείριση των ερευνητικών έργων με ψηφιοποιημένο υλικό και συλλογών με εικόνες. Το Preservation Metadata for Digital Collection έχει βασιστεί σε πολλές αντίστοιχες προσπάθειες, μερικές εκ των οποίων είναι το Open Archive Information System (OAIS), το Networked European Deposit Library (NEDLIB), το Curl Exemplars in Digital Archives (CEDARS), κ.ά.

Αποτελείται από 25 στοιχεία τα οποία είναι:

- Μόνιμος προσδιοριστής (Persistent Identifier).
- Ημερομηνία δημιουργίας (Date of Creation).
- Είδος δομής (Structural Type).
- Τεχνική υποδομή (Technical Infrastructure of Complex Object).
- Περιγραφή αρχείου (File Description).
- Συνήθειες απαιτήσεις συστήματος (Known System Requirements).
- Απαιτήσεις εγκατάστασης (Installation Requirements).

- Πληροφορίες αποθήκευσης (Storage Information).
- Απαγορεύσεις πρόσβασης (Access Inhibitors).
- Αναζήτηση και πρόσβαση (Finding & Searching Aids, and Access Facilitators).
- Άδεια πράξης διατήρησης (Preservation Action Permission).
- Εγκυρότητα (Validation).
- Σχέσεις (Relationships).
- Ιδιαιτερότητες (Quirks).
- Απόφαση αρχειακής εργασίας (Archiving Decision-work).
- Αιτιολόγηση απόφασης εργασίας (Decision Reason-work).
- Υπεύθυνος θεσμός για αρχειοθέτηση (Institution Responsible for Archiving Decision-work).
- Δήλωση αρχειακής απόφασης (Archiving Decision-manifestation).
- Δήλωση αιτιολόγησης της απόφασης (Decision Reason-manifestation).
- Υπεύθυνος θεσμός για την πολιτική αρχειοθέτησης (Institution Responsible for Archiving Decision-manifestation).
- Είδος σκοπού (Intention Type).
- Υπεύθυνος θεσμός για τη διατήρηση (Institution with preservation responsibility).
- Επεξεργασία (Process).
- Δημιουργός εγγραφής (Record Creator).
- Άλλα (Other).

Σημειώνεται ότι το PANDORA μαζί με το CEDARS<sup>14</sup> προσπαθούν να αναπτύξουν συνεργατικά ένα διεθνές πρότυπο

13. Lynch, C., 1999. Building Digital Libraries for Metadiversity: Federation Across Disciplines, *Proceedings of the Symposium Sponsored by the U.S. Geological Survey, Biological Resources Division & the National Federation of Abstracting and Information Services.*

14. Law, C., 2000. PANDORA-Towards a National Collection of Selected Australian Online Publications. *66th IFLA Council and General Conference, Jerusalem, Israel, 13-18 August (Ημερομηνία πρόσβασης: 17.01.05)*

μεταδεδομένων διατήρησης.

### Στατιστικά

Τα στατιστικά<sup>15</sup> αναφέρονται στο μέγεθος του αρχείου και στη μηνιαία ανάπτυξη του όπως αυτά προέκυψαν στις 15 Δεκεμβρίου 2004.

	Νοέμβριος	Δεκέμβριος	Ανάπτυξη
Συνολικός αριθμός αρχειοθετημένων τίτλων	7,404	7,527	147
Συνολικός αριθμός αρχειοθετημένων τεκμηρίων (archived instances)	14,955	14,687	268
Συνολικός αριθμός αρχείων	22,6637996	22,491,536	172,263
Συνολικό μέγεθος σε Gigabits	760,721	753,4	7,321

Ένα αρχειοθετημένο τεκμήριο (archived instance) είναι ένα στιγμιότυπο (snapshot) ή αντίτυπο ενός τίτλου που έχει προστεθεί στο αρχείο. Πολλοί τίτλοι αντιγράφονται στο αρχείο περισσότερο από μια φορά για να δηλώσουν τις αλλαγές στο περιεχόμενο, π.χ. περιοδικοί τίτλοι που προσθέτουν νέα τεύχη. Κάθε νέο αντίγραφο ενός τίτλου δηλώνεται ως 'archived instance'. Παρατηρούμε ότι η μηνιαία ανάπτυξη των αρχειοθετημένων τίτλων ανέρχεται σε 147 τίτλους, ενώ ο συνολικός αριθμός αρχειοθετημένων τεκμηρίων έχει μια αύξηση της τάξεως των 268 νέων τεκμηρίων.

### Συμπεράσματα

Το PANDORA αποτελεί μια από τις ση-

μαντικότερες διεθνείς πρωτοβουλίες για τη διατήρηση των ψηφιακών τεκμηρίων με σκοπό τη μακροβιότητα της αυστραλιανής πολιτιστικής κληρονομιάς. Η Εθνική Βιβλιοθήκη της Αυστραλίας ξεκινώντας πολύ νωρίς, για τα ελληνικά δεδομένα, κατάφερε να πρωτοπορήσει στην αρχειοθέτηση του Ιστού και να αναπτύξει το PANDORA προβλέποντας όλες τις λεπτομέρειες και τις διαδικασίες. Το PANDORA είναι ένα αρχείο προς μίμηση, αποτελώντας παράδειγμα για τις υπόλοιπες Εθνικές Βιβλιοθήκες για την αρτιότητα και τον προγραμματισμό του. Είναι ένα Web αρχείο ελεύθερα προσβάσιμο, λειτουργώντας ως ένα ακόμα εργαλείο έρευνας για τους δυνητικούς χρήστες.

### Βιβλιογραφία

1. Brandis, L., Lyall J., 2003. PADI: Preserving Access to Australian Information and Cultural Heritage in Digital Form. Διαθέσιμο στη διεύθυνση: <http://www.nla.gov.au/nla/staffpaper/lyall3.html> (Ημερομηνία πρόσβασης: 17.01.05)
2. Hodge, G.M., 2000. Best Practices for Digital Archiving: An Information Life Cycle. In: *D-Lib Magazine*, vol. 6, no. 1. Διαθέσιμο στη διεύθυνση: <http://www.dlib.org/dlib/january00/01hodge.html> (Ημερομηνία πρόσβασης: 19.01.05)
3. Law, C., 2000. PANDORA-Towards a National Collection of Selected Australian Online Publications. 66th *IFLA Council and General Conference*,

15. NLA, 2004. Pandora Statistics. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/statistics.html> (Ημερομηνία πρόσβασης: 19.01.05)

Jerusalem, Israel, 13-18 August (Ημερομηνία πρόσβασης: 17.01.05)

**4.** Lynch, C., 1999. Building Digital Libraries for Metadiversity: Federation Across Disciplines. *Proceedings of the Symposium Sponsored by the U.S. Geological Survey, Biological Resources Division & the National Federation of Abstracting and Information Services*

**5.** NLA, 2003. PANDORA Business Process Model. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/bpm.html> (Ημερομηνία πρόσβασης: 17.01.05)

**6.** NLA, 2003. PANDORA. Australia's Web Archive: an Overview. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/background.html> (Ημερομηνία πρόσβασης: 19.01.05)

**7.** NLA, 2003. Business Process Model: 1.3 Online Electronic Publications. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/bpm.html#onli> (Ημερομηνία πρόσβασης: 17.01.05)

**8.** NLA, 2003. Online Australia Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/selectionguidelines.html> (Ημερομηνία πρόσβασης: 17.01.05)

**9.** NLA, 2003. PANDORA Business Process Model: 4. The archive management component. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/bpm.html> (Ημερομηνία πρόσβασης: 16.01.05)

**10.** NLA, 2003. PANDORA Business

Process Model: 5. Issues of Archive Management. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/bpm.html> (Ημερομηνία πρόσβασης: 17.01.05)

**11.** NLA, 2004. Pandora Statistics. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/statistics.html> (Ημερομηνία πρόσβασης: 19.01.05)

**12.** NLA, 1999. Preservation Metadata for Digital Collections. Διαθέσιμο στη διεύθυνση: <http://www.nla.gov.au/preserve/pmeta.html> (Ημερομηνία πρόσβασης: 17.01.05)

**13.** PANDORA. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/> (Ημερομηνία πρόσβασης: 17.01.05)

**14.** Phillips, M.E., 2003. Collecting Australian Online Publications-Balanced Scorecard Initiative 49. Διαθέσιμο στη διεύθυνση: <http://pandora.nla.gov.au/BSC49.doc> (Ημερομηνία πρόσβασης: 16.01.05)

**15.** SCOAP. Διαθέσιμο στη διεύθυνση: <http://www.nla.gov.au/scoap/> (Ημερομηνία πρόσβασης: 19.01.05)

*Ο Διονύσης Κόκκινος είναι βιβλιοθηκονόμος στο Εθνικό Μετσόβιο Πολυτεχνείο και Γενικός Γραμματέας του Δ.Σ. της Ε.Ε.Β.Ε.Π.*

*Η Μαρία Γαβρίλη είναι βιβλιοθηκονόμος στο Athens Information Technology.*