

ΠΟΙΟΤΗΤΑ ΒΙΒΛΙΟΓΡΑΦΙΚΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΣΤΟΧΟ ΤΗ ΔΙΑΛΕΙΤΟΥΡΓΙΚΟΤΗΤΑ: Η ΠΕΡΙΠΤΩΣΗ ΤΗΣ ΕΘΝΙΚΗΣ ΒΙΒΛΙΟΘΗΚΗΣ ΤΗΣ ΕΛΛΑΔΟΣ

Γερόλιμος, Μ., Ζαπουνίδου, Σ., Κουφάκου, Ε.

Εθνική Βιβλιοθήκη της Ελλάδος

Λέξεις κλειδιά: Ποιότητα δεδομένων, Έλεγχος ποιότητας, Διαλειτουργικότητα, Βιβλιογραφικά δεδομένα, Καθιερωμένοι όροι

Ο ρόλος της Εθνικής Βιβλιοθήκης της Ελλάδος (ΕΒΕ) ως Εθνικό Βιβλιογραφικό και Καταλογογραφικό Κέντρο (ν. 4452/2017) είναι η έγκαιρη και ολοκληρωμένη καταγραφή της τρέχουσας πνευματικής παραγωγής. Το 2022 η ΕΒΕ προχώρησε σε δύο δομικές αλλαγές στη διαχείριση των βιβλιογραφικών δεδομένων που αυτή παράγει. Αυτές ήταν η μετάβαση από το UNIMARC στο MARC21 ως πρότυπο διάταξης δεδομένων, και η υιοθέτηση των κανόνων βιβλιογραφικής επεξεργασίας του επίσημου Resource Description and Access (RDA), σε αντικατάσταση των ΑγγλοΑμερικάνικων Κανόνων Καταλογογράφησης (AACR2). Το επίσημο RDA αποτελεί ένα πρότυπο συμβατό με τις τεχνολογίες συνδεδεμένων δεδομένων (linked data) και προωθεί τη χρήση ελεγχόμενων λεξιλογίων και δομημένης πληροφορίας για λόγους διαλειτουργικότητας και διαμοιρασμού των δεδομένων. Σε αυτό το πλαίσιο, το Τμήμα Καταλόγων της ΕΒΕ υιοθέτησε και εφάρμοσε μία σειρά εργασιών και ενεργειών για τον μαζικό έλεγχο της ποιότητας των δεδομένων που παράγει με στόχο τη συνεχή αναβάθμιση της ποιότητας των εγγραφών του καταλόγου, πριν, κατά τη διάρκεια και μετά την παραδοσιακή εργασία της καταλογογράφησης.

Στην εργασία, θα παρουσιαστούν α) οι έννοιες της ποιότητας δεδομένων και των σχετικών παραμέτρων, και β) η σχεδίαση και υλοποίηση της εργασίας ελέγχου ποιότητας στα βιβλιογραφικά δεδομένα της ΕΒΕ. Η ροή εργασίας που έχει δημιουργηθεί περιλαμβάνει, εκτός από τα προφανή στάδια ανίχνευσης και διόρθωσης λαθών, ενέργειες συστηματικού εμπλουτισμού παλαιών και νέων δεδομένων είτε με δομημένη πληροφορία, είτε με τιμές από ελεγχόμενα λεξιλόγια. Εκτός από τις μεθοδολογικές επιλογές, θα παρουσιαστούν στατιστικά δεδομένα από τις σχετικές εργασίες που πραγματοποιούνται από τους βιβλιοθηκονόμους συστημάτων τα τελευταία δυο χρόνια.

Στόχος της παρουσίασης είναι να παρουσιαστεί στην ελληνική βιβλιοθηκονομική κοινότητα η πρακτική της ΕΒΕ στη διαχείριση των βιβλιογραφικών της δεδομένων και να αποτελέσει αντικείμενο εξωστρέφειας και, ενδεχομένως, χρήσης από άλλες βιβλιοθήκες, με στόχο την ανάπτυξη και υιοθέτηση κοινών εργαλείων που βελτιώνουν, όχι μόνο την ποιότητα των δεδομένων των καταλόγων των βιβλιοθηκών, αλλά και το πλαίσιο συνεργασίας μεταξύ των ελληνικών βιβλιοθηκών, όπως συνειδητά επιδιώκει τα τελευταία χρόνια η ΕΒΕ.

QUALITY OF BIBLIOGRAPHIC DATA FOR INTEROPERABILITY: THE CASE OF THE NATIONAL LIBRARY OF GREECE

Gerolimos, M., Zapounidou, S., Koufakou, E
National Library of Greece

Keywords: Data quality, Quality control, Interoperability, Bibliographic data, Authorities

The role of the National Library of Greece (NLG) as the National Bibliographic and Cataloguing Centre is the timely and comprehensive recording of current intellectual production. In 2022, the NLG made two structural changes in the management of the bibliographic data it produces. These were the transition from UNIMARC to MARC21 as the data communication format, and the adoption of the official Resource Description and Access (RDA) rules, replacing the Anglo-American Cataloguing Rules (AACR2). The official RDA is a standard compatible with linked data technologies. For interoperability and data sharing purposes, it promotes the use of controlled vocabularies and structured information. In this context, the NLG has developed a series of tasks and actions for checking the data produced by the Cataloguing Department.

In the paper, we will present a) the concepts of data quality and its dimensions, and b) the design and implementation of the quality control workflow at the NLG. The workflow includes, in addition to the obvious error detection and correction steps, actions for the systematic enrichment of legacy and new data with either structured information or values from controlled vocabularies. In addition to the methodological choices, statistical data from related work carried out by systems librarians over the last two years will be presented.

The aim of the presentation is to introduce the NLG quality control workflow to the Greek library community hoping that the net result would be the development and adoption of common tools that will improve the quality of library data, as well as provide a framework for cooperation between Greek libraries.

Εισαγωγή

Η παραδοσιακή παραγωγή βιβλιογραφικών δεδομένων στηρίχθηκε στην καταγραφή πληροφοριών σε εγγραφές. Το κύριο αντικείμενο περιγραφής υπήρξαν οι εκδόσεις και με αφορμή αυτές (και ανάλογα με την πολιτική κάθε βιβλιοθήκης) καταγράφονταν επιπλέον πληροφορίες για σχετιζόμενες οντότητες όπως συγγραφείς, εκδότες, θέματα, γεωγραφικά ονόματα, κ.λπ. Ωστόσο, το ακριβές νόημα των σχετιζόμενων οντοτήτων, ο ορισμός των απαραίτητων πληροφοριών που πρέπει να περιλαμβάνουν οι περιγραφές τους και η πλαισίωσή τους με διασυνδέσεις προς άλλες οντότητες δεν ήταν ξεκάθαρα. Το μοντέλο που σήμερα παρέχει το πλαίσιο εργασίας στη βιβλιοθηκονομική κοινότητα είναι το IFLA Library Reference Model (LRM). Ορίζει τις οντότητες ενδιαφέροντος, τις ιδιότητές τους και τις μεταξύ τους σχέσεις. Το μοντέλο Resource Description and Access, γνωστό ως RDA, βασίζεται στο LRM και το επεκτείνει ορίζοντας επιπλέον οντότητες, ιδιότητες και σχέσεις. Το RDA είναι συμβατό με τις τρέχουσες τεχνολογίες απόδοσης σημασιολογίας και διασύνδεσης, γνωστές ως Σημασιολογικός Ιστός και Συνδεδεμένα Δεδομένα (Semantic Web και Linked Data).

Η Εθνική Βιβλιοθήκη της Ελλάδος (ΕΒΕ) έχει επιλέξει την εφαρμογή των νέων μοντέλων για τα δεδομένα που παράγει. Η απόφαση αυτή έχει επιφέρει σημαντικές αλλαγές στη διαμόρφωση της πολιτικής καταλογογράφησης και στον τρόπο επεξεργασίας των δεδομένων. Στο πλαίσιο αυτό, έχουν επιλεγεί οι οντότητες ενδιαφέροντος, ο τρόπος περιγραφής τους, τα ελεγχόμενα λεξιλόγια και οι συσχετίσεις. Η εφαρμογή και ο έλεγχος των επιλογών αυτών πραγματοποιείται μέσα από την καθημερινή ροή εργασίας, ενώ, αξιοποιώντας τις δυνατότητες των νέων μοντέλων, η ΕΒΕ μετατρέπει, πιλοτικά, μέρος των δεδομένων του καταλόγου σε συνδεδεμένα δεδομένα. Για τη μετατροπή όλων των δεδομένων της ΕΒΕ οι εργασίες ελέγχου επεκτάθηκαν και στα παλαιότερα δεδομένα της ΕΒΕ τα οποία δημιουργήθηκαν βάσει προηγούμενων κανόνων καταλογογράφησης, από διαφορετικό κατά εποχές προσωπικό και κατ' εφαρμογή διαφορετικών διαδικασιών.

Η εργασία παρουσιάζει τη ροή εργασιών που αναπτύχθηκε και εφαρμόζεται από το Τμήμα Καταλόγων της ΕΒΕ για τη δημιουργία περιγραφών που να ανταποκρίνονται στις απαιτήσεις της πολιτικής της με έμφαση στην ποιότητα, τη διαλειτουργικότητα και την υποστήριξη πολυγλωσσίας.

Βιβλιογραφική ανασκόπηση

Οι αρχές, τα μοντέλα και οι κανόνες βιβλιογραφικής επεξεργασίας παρέχουν ένα πλαίσιο δημιουργίας περιγραφών για οντότητες βιβλιογραφικού ενδιαφέροντος. Στην πράξη, όμως, η δημιουργία των περιγραφών ποικίλλει, καθώς παρουσιάζονται συνεχώς νέες μορφές υλικού με νέες περιγραφικές απαιτήσεις. Επιπλέον, τα δεδομένα των βιβλιοθηκών έχουν δημιουργηθεί σε πολύ διαφορετικές συνθήκες. Κάποια παρέχονται από εκδότες, άλλα από τους ίδιους τους δημιουργούς με ελάχιστη τεκμηρίωση, άλλα κατ' εφαρμογή παλαιών τοπικών κανόνων ή/και πολιτικών. Αυτό που προκύπτει είναι κατάλογοι βιβλιοθηκών με ετερογενείς εγγραφές που παρουσιάζουν ποικιλία σε πλήθος μεταδεδομένων, διαφορετικά επίπεδα περιγραφής και ποιότητας.

Εργασίες ελέγχου ποιότητας εκτελούνται κατά καιρούς σε όλες τις βιβλιοθήκες ακόμα και αν δεν έχει εφαρμοστεί μία τακτική διαδικασία. Η εμπειρία δείχνει πως η συμμετοχή σε συλλογικούς καταλόγους, ο ενεργός διαμοιρασμός δεδομένων με τρίτες βιβλιοθήκες (π.χ. αποστολή δεδομένων

στο VIAF) και η υλοποίηση έργων αλλαγής λογισμικού έχουν επιφέρει αναγκαστικούς ελέγχους ποιότητας των δεδομένων (Snow, 2017). Ωστόσο, οι βελτιώσεις που προκύπτουν από τους ελέγχους αυτούς, αφορούν κυρίως την μερική βελτίωση των μεταδεδομένων για την ομαλή μεταφορά τους από σύστημα σε σύστημα και δεν στοχεύουν στην βελτίωση της ποιότητας συνολικά. Αν ληφθεί υπόψη ότι, τα μεταδεδομένα εξελίσσονται (ή πρέπει να εξελίσσονται) ακολουθώντας τον κύκλο ζωής των αντικειμένων ή των οντοτήτων που περιγράφουν (Gilliland, 2016), η εφαρμογή μίας συστηματικής διαδικασίας ελέγχου και εμπλουτισμού των μεταδεδομένων κρίνεται άκρως απαραίτητη.

Το βασικό ερώτημα που προκύπτει είναι: τι ορίζεται ως ποιότητα; Η ποιότητα είναι μία σχετική έννοια, μπορεί να λάβει διαφορετική ερμηνεία ανάλογα με την οπτική προσέγγισης ή το κριτήριο αξιολόγησης. Για κάποια βιβλιοθήκη η ύπαρξη συγκεκριμένων πληροφοριών σε εγγραφές μπορεί να αποτελεί απόδειξη ποιότητας, ενώ για κάποια άλλη οι ίδιες εγγραφές μπορεί να κριθούν अपαράδεκτες καθώς απουσιάζει η ISBD στίξη (Boissonnas, 1979). Λαμβάνοντας ως δεδομένο ότι δεν υπάρχει «τέλεια εγγραφή», η επιδίωξη των βιβλιοθηκών πρέπει να στραφεί προς μία «έντιμη λύση» (Bade, 2008)· την παραγωγή εγγραφών που όταν αξιολογούνται πληρούν συγκεκριμένες απαιτήσεις.

Το πρώτο βήμα σε έναν έλεγχο ποιότητας είναι ο ορισμός των απαιτήσεων ή διαφορετικά η οπτική προσέγγισης. Η βιβλιογραφική έρευνα της Badovinac (2017, 2021) κατέγραψε 5 διαφορετικές οπτικές προσέγγισης της διαδικασίας ελέγχου ποιότητας στην καταλογογράφηση:

- **Τελικοί χρήστες.** Λαμβάνει υπόψη κυρίως τις απαιτήσεις και τις προσδοκίες των χρηστών με αποτέλεσμα ο έλεγχος να διεξάγεται μέσω ποιοτικών ερευνών χρηστών.
- **Ροή εργασιών.** Εστιάζει στη διοικητική διαχείριση των εργασιών δημιουργίας και αναθεώρησης μεταδεδομένων, λαμβάνοντας υπόψη δείκτες παραγωγικότητας και κόστους.
- **Καταλογογράφοι και ειδικοί μεταδεδομένων.** Επικεντρώνεται στην ποιότητα των μεταδεδομένων σε συνδυασμό με τις γνώσεις, τα εργαλεία και τα εγχειρίδια που απαιτούνται για τη βελτίωσή τους. Οι έλεγχοι ποιότητας αυτού του τύπου συχνά συνδυάζονται με συνεντεύξεις στο προσωπικό και ποιοτικές έρευνες για ανίχνευση παραγόντων που μπορούν να βελτιώσουν την ποιότητα των μεταδεδομένων (όπως αυτή ορίζεται από τον εκάστοτε φορέα).
- **Εργαλεία καταλογογράφησης.** Συναντάται, συνήθως, σε έργα μετάπτωσης από ένα σύστημα σε ένα άλλο με γνώμονα την ομαλή μεταφορά των μεταδεδομένων.
- **Προτυποποίηση.** Στοχεύει να εξετάσει το κατά πόσο τηρούνται οι κανόνες των προτύπων σε σχέση πάντα με την τοπική πολιτική εφαρμογής. Αποτελεί ίσως την πιο διαδεδομένη προσέγγιση ελέγχου ποιότητας.

Το επόμενο βήμα είναι ο καθορισμός της μεθόδου ελέγχου και των κριτηρίων που θα χρησιμοποιηθούν. Αν και πολλοί ερευνητές έχουν μελετήσει το πεδίο της ποιότητας μεταδεδομένων αποσκοπώντας στον ορισμό συγκεκριμένων κριτηρίων (βλ. Kumar, 2024), η εργασία λαμβάνει υπόψη τη θεμελιώδη έρευνα των Bruce και Hillmann (2004) η οποία ορίζει 7 κριτήρια ποιότητας.

- **Πληρότητα (completeness).** Η κάθε περιγραφή αναπαριστά πλήρως το αντικείμενό της. Καταγράφονται όλες οι απαιτούμενες πληροφορίες. Προϋποθέτει επιπλέον, ότι όλες οι περιγραφές με το ίδιο αντικείμενο (π.χ. εγγραφές περιγραφής φυσικών προσώπων) περιλαμβάνουν τις απαιτούμενες πληροφορίες.
- **Ακρίβεια (accuracy).** Η κάθε περιγραφή περιλαμβάνει ακριβείς πληροφορίες και σχετική τεκμηρίωση. Στην παράμετρο αυτή περιλαμβάνεται η αποφυγή ορθογραφικών και συντακτικών λαθών και η χρήση κανόνων στην καταγραφή της τεκμηρίωσης.
- **Προέλευση (provenance).** Η κάθε περιγραφή ενημερώνει για την προέλευση των πληροφοριών που παρέχει ή περιλαμβάνει. Χαρακτηριστικά παραδείγματα είναι η καταγραφή της βιβλιοθήκης που δημιούργησε ή εμπλούτισε μία περιγραφή, το επίπεδο περιγραφής, αν κάποια πληροφορία έχει προστεθεί με αυτοματοποιημένο τρόπο, κ.λπ.
- **Συμμόρφωση στις προσδοκίες (conformance to expectations).** Η κάθε περιγραφή περιλαμβάνει τις πληροφορίες που προβλέπονται από το προφίλ εφαρμογής και όλα τα προβλεπόμενα πεδία έχουν συμπληρωθεί με τρόπο που να ικανοποιούν τις προσδοκίες της ομάδας-στόχου (π.χ. τελικοί χρήστες, βιβλιοθηκονόμοι).
- **Λογική συνέπεια και συνοχή (logical consistency and coherence).** Η κάθε περιγραφή (ανεξάρτητα από την προέλευση ή το πλαίσιο δημιουργίας της) περιλαμβάνει όλες τις απαιτούμενες από την πολιτική της βιβλιοθήκης πληροφορίες. Η αξιοποίηση όλων των περιγραφών κατά την αναζήτηση είναι δείγμα λογικής συνέπειας και συνοχής ενός καταλόγου.
- **Επικαιρότητα (timeliness).** Η κάθε περιγραφή διατίθεται έγκαιρα και ανανεώνεται ανάλογα με το αν υπάρχει κάποια αλλαγή στο αντικείμενο ή οντότητα περιγραφής.
- **Προσβασιμότητα (accessibility).** Τα δεδομένα είναι κατανοητά και προσβάσιμα σε όλους τους χρήστες. Εδώ θα μπορούσε να προστεθεί και η έννοια της διαλειτουργικότητας με την έννοια της προσβασιμότητας σε μηχανικούς χρήστες.

Φυσικά ανάλογα με την πολιτική και τις δυνατότητες σε υλικοτεχνικές υποδομές και προσωπικό μίας βιβλιοθήκης κάποιες παράμετροι μπορεί να κριθούν περισσότερο σημαντικές από άλλες. Για κάθε παράμετρο χρειάζεται να οριστούν οι τρόποι ελέγχου και η συχνότητα που αυτοί θα εκτελούνται. Επίσης, πρέπει να σχεδιαστεί και ο τρόπος αντιμετώπισης πιθανών προβλημάτων. Η επόμενη ενότητα παρουσιάζει την προσέγγιση της ΕΒΕ στον έλεγχο ποιότητας καθιερωμένων και βιβλιογραφικών εγγραφών.

Έλεγχος ποιότητας δεδομένων

Η ΕΒΕ έχει θέσει ως στόχο την εφαρμογή των νέων μοντέλων βιβλιογραφικής επεξεργασίας με τρόπο που να καθιστά τα μεταδεδομένα της έτοιμα για εφαρμογές συνδεδεμένων δεδομένων (linked-data ready). Ο έλεγχος ποιότητας προσεγγίζεται με έμφαση στην προτυποποίηση βάσει των προφίλ εφαρμογής της ΕΒΕ και στις ροές εργασίας που θα διασφαλίζουν την παραγωγή ποιοτικών μεταδεδομένων.

Σημείο έναρξης αποτελούν τα προφίλ εφαρμογής που έχει δημιουργήσει το Τμήμα Καταλόγων για την περιγραφή οντοτήτων σύμφωνα με το επίσημο RDA. Μέχρι στιγμής έχουν διαμορφωθεί προφίλ

για τις οντότητες Φυσικό πρόσωπο (Person), Όνομα (Nomen), και Εκδήλωση (Manifestation). Έχει, επίσης, δημιουργηθεί για την οντότητα Έννοια (SKOS Concept), η οποία δεν ανήκει στις οντότητες του IFLA LRM και του RDA, αλλά χρησιμοποιείται για την περιγραφή εννοιών και θεμάτων. Τα προφίλ εφαρμογής, επί της ουσίας, ορίζουν τα MARC πεδία και υποπεδία που πρέπει να περιλαμβάνονται σε μία εγγραφή και τον τρόπο με τον οποίο συμπληρώνονται, π.χ., αδόμητη περιγραφή με ελεύθερο κείμενο ή δομημένη περιγραφή με χρήση ελεγχόμενων λεξιλογίων.

Ακολουθεί ο έλεγχος εφαρμογής με στόχο την αποφυγή λαθών που πρέπει να εντοπιστούν και να διορθωθούν στη συνέχεια. Για την επίτευξη αυτού του στόχου, ανασχεδιάστηκε, το 2022, η ροή εργασίας του Τμήματος Καταλόγων, όπως περιγράφεται στη συνέχεια.

Ανασχεδιασμός ροής εργασίας

Από το 2014 η ΕΒΕ χρησιμοποιεί το βιβλιοθηκονομικό σύστημα Koha, το οποίο παρέχει τη δυνατότητα διαμόρφωσης πλαισίων βιβλιογραφικών και καθιερωμένων εγγραφών. Τα πλαίσια αυτά ενημερώθηκαν βάσει των προφίλ εφαρμογής και εμπλουτίστηκαν με ελεγχόμενα λεξιλόγια και κανόνες, με στόχο αφενός τη διευκόλυνση των καταλογογράφων στην καθημερινή εργασία και αφετέρου την αποφυγή λαθών κατά την πληκτρολόγηση.

Λόγω της λειτουργίας του Κέντρου ISBN-ISSN-ISMN, αλλά και της Κατά Νόμο Κατάθεσης, το Τμήμα Καταλόγων παραλαμβάνει βιβλιογραφικές εγγραφές προς εισαγωγή στον Κατάλογο για τον μεγαλύτερο αριθμό τεκμηρίων της συλλογής της ΕΒΕ. Οι εγγραφές αυτές έχουν συμπληρωμένα βασικά πεδία περιγραφής, όπως τίτλος, συγγραφέας, εκδότης κ.λπ. Από το 2022 εφαρμόζεται η επεξεργασία των εγγραφών πριν εισαχθούν στον κατάλογο, με τη μαζική εισαγωγή τιμών (π.χ. στα πεδία 3xx), τη διόρθωση (π.χ. συντομογραφιών) και τον εμπλουτισμό (π.χ. τιμές σε κωδικοποιημένα πεδία), μειώνοντας τον χρόνο καταλογογράφησης ενός τεκμηρίου και αυξάνοντας την ποιότητα των εγγραφών.

Μετά την καταλογογράφηση των τεκμηρίων, η ομάδα βιβλιοθηκονόμων συστημάτων ελέγχει τις εγγραφές, προχωρά σε διορθώσεις και εμπλουτισμούς (π.χ. προσθήκη URIs) και ταυτόχρονα ανιχνεύει επιπλέον σημεία ελέγχου σε νέες και υπάρχουσες εγγραφές για μελλοντικές διορθώσεις και εμπλουτισμούς.

Σχεδιασμός διαδικασίας ελέγχου

Ο σχεδιασμός της διαδικασίας ελέγχου λαμβάνει υπόψη όλα τα κριτήρια των Bruce και Hillmann (2004) με αποτέλεσμα να δημιουργηθεί μία λίστα με πεδία προς έλεγχο για κάθε RDA οντότητα. Ο Πίνακας 1 παρουσιάζει ένα σχετικό παράδειγμα για την οντότητα RDA Person – Φυσικό πρόσωπο.

Πίνακας 1. Καθορισμός πεδίων ελέγχου για την οντότητα RDA Φυσικό Πρόσωπο με βάση τα 7 κριτήρια ποιότητας των Bruce & Hillman (2004)

Κριτήριο	Φυσικό πρόσωπο
Πληρότητα	πεδία προφίλ εφαρμογής
Ακρίβεια	ορθογραφικός έλεγχος σε 6XX έλεγχος τιμών σε 3XX
Προέλευση	έλεγχος τιμών σε 040 και σε 4XX\$7
Συμμόρφωση με προσδοκίες	πεδία προφίλ εφαρμογής
Λογική συνέπεια και συνοχή	πεδία προφίλ εφαρμογής έλεγχος τιμών σε 0XX, 3XX, 4XX\$i\$4, 5XX\$i\$4, 6XX έλεγχος δεικτών σε 016, 024, 100, 377, 4X0, 5X0
Επικαιρότητα	Έτη γέννησης και θανάτου
Προσβασιμότητα	Χρήση URIs

Ο έλεγχος πραγματοποιείται με εισαγωγή εγγραφών στο MarcEdit και εξαγωγή πεδίων/υποπεδίων σε φύλλα εργασίας. Η επιλογή των πεδίων και υποπεδίων πραγματοποιείται βάσει του προφίλ εφαρμογής της κάθε οντότητας και τον πίνακα κριτηρίων. Ο Πίνακας 2 παρουσιάζει ένα μέρος των πεδίων και υποπεδίων που εξάγονται για την οντότητα Φυσικό πρόσωπο και των ελέγχων που διενεργούνται σε αυτά.

Πίνακας 2. Τμήμα του πίνακα ελέγχου για την οντότητα Φυσικό Πρόσωπο με επιλογές πεδίων/υποπεδίων, τύπο ελέγχου και αναμενόμενες τιμές.

Field	Subfield	Expected values	Description of checkpoint	ME_Person_0xx	ME_Person_0xxx	ME_Person_3xx	ME_Person_4xx_5x
24	1	RWO URI	έλεγχος για αναντιστοιχία επιλογών στα 3 υποπεδία του 024 έλεγχος για κενά έλεγχος για σωστή μορφή URI		Y		
40	a	GRATEVE	έλεγχος για μη αναμενόμενες τιμές	Y			
40	b	gre	έλεγχος για μη αναμενόμενες τιμές	Y			
40	c	GRATEVE	έλεγχος για μη αναμενόμενες τιμές	Y			
40	d	GRATEVE	να ελεγχθεί σε συνδυασμό με το 977\$b modified	Y			
40	e	rda	να ελεγχθεί σε συνδυασμό με το 075\$a	Y	Y	Y	Y
40	f	nlgaF	έλεγχος για μη αναμενόμενες τιμές	Y			
46	f	date	έλεγχος για αναντιστοιχία επιλογών στα υποπεδία του 046 Αντιγράφο σε νέο φύλλο 001, 040\$e, 046\$f, 046\$g, 100\$d Κανονικοποίηση δεδομένων (βλ. παρακάτω)		Y		
46	g	date	έλεγχος για αναντιστοιχία επιλογών στα υποπεδία του 046		Y		
46	s	date or century	έλεγχος για αναντιστοιχία επιλογών στα υποπεδία του 046		Y		
46	t	date or century	έλεγχος για αναντιστοιχία επιλογών στα υποπεδία του 046		Y		
46	2	edtf or blank	έλεγχος για αναντιστοιχία επιλογών στα υποπεδία του 046		Y		
75	a		έλεγχος για αντιστοιχία σε σχέση με 100 \$a, \$t, \$l, \$v, \$x, \$y, \$z	Y	Y	Y	Y
75	b		έλεγχος για αντιστοιχία σε σχέση με 100 \$a, \$t, \$l, \$v, \$x, \$y, \$z		Y		
75	2		έλεγχος για αντιστοιχία σε σχέση με 100 \$a, \$t, \$l, \$v, \$x, \$y, \$z		Y		
75	1		έλεγχος για αντιστοιχία σε σχέση με 100 \$a, \$t, \$l, \$v, \$x, \$y, \$z		Y		
100	a		σε συνδυασμό με 100 \$d		Y	Y	Y
100	b		σε συνδυασμό με 100 \$d		Y		
100	q		σε συνδυασμό με 100 \$d		Y		
100	c		σε συνδυασμό με 100 \$d		Y		
100	d		έλεγχος για κόμμα πριν το \$d Έλεγχος για 046		Y		Y

Η ανίχνευση λαθών γίνεται με αξιοποίηση φίλτρων και συναρτήσεων στο MS Excel. Οι διορθώσεις, ανάλογα με τη φύση του λάθους, γίνονται είτε μαζικά στο Κοha, είτε χειροκίνητα, είτε, σπανιότερα, με αντικατάσταση των εγγραφών μετά τον εμπλουτισμό τους. Αξίζει να σημειωθεί ότι, κατά την ανίχνευση λαθών εντοπίζονται και ευκαιρίες διορθώσεων ή εμπλουτισμού παλαιότερων εγγραφών.

Εκτέλεση ελέγχων και αποτελέσματα

Η συχνότητα εκτέλεσης των ελέγχων είναι διαφορετική για κάθε οντότητα. Για παράδειγμα, ο έλεγχος των εγγραφών που περιγράφουν Φυσικά Πρόσωπα (RDA Person) και Ονόματα/Ψευδώνυμα (RDA Nomen) γίνεται ανά 15νθήμερο, ενώ οι εγγραφές που περιγράφουν SKOS Έννοιες - Θέματα

ελέγχονται σε μηνιαία βάση. Αν και έχουν γίνει έλεγχοι σε εγγραφές της οντότητας Εκδήλωση (RDA Manifestation), οι έλεγχοι αυτοί δεν έχουν ακόμα συστηματοποιηθεί.

Οι πρώτοι έλεγχοι επικεντρώθηκαν στη διόρθωση λαθών σε κωδικοποιημένα πεδία και στη διόρθωση δεικτών. Ο Πίνακας 3 παρουσιάζει ένα τμήμα των αλλαγών στις καθιερωμένες εγγραφές με βασικό πεδίο το 100 που περιγράφουν Φυσικά Πρόσωπα και Ονόματα/Ψευδώνυμα. Για παράδειγμα, μία καταλογογράφος επιλέγει ως Σχετική γλώσσα ενός φυσικού προσώπου τον κωδικό gre στο 377\$a. Κατά τον έλεγχο της εγγραφής η ομάδα συστημάτων προσθέτει μαζικά τα υπόλοιπα υποπεδία \$1, \$2, και \$1. Επίσης, οι εγγραφές εμπλουτίζονται με επιπλέον πληροφορίες που θα εξυπηρετήσουν σε καινοτόμες προβολές αυτών των δεδομένων στο μέλλον, όπως, ενδεικτικά, τα κρατικά βραβεία λογοτεχνίας, τα λογοτεχνικά ρεύματα και οι σχέσεις με άλλα φυσικά πρόσωπα και συλλογικά όργανα.

Πίνακας 3. Τμήμα του πίνακα παρακολούθησης αλλαγών και εμπλουτισμών σε εγγραφές φυσικών προσώπων και ψευδωνύμων

Ημερομηνία dataset	Έλεγχος	ME Person_0xx	ME Person_0xx+	ME Person_3xx	ME Person_4xx_5xx	Επιπλέον Έλεγχοι	Indicators	Σύνολο διορθώσεων	Παρατηρήσεις
28/3/2023	29/3/2023	252	85	282	78	682	0	1.379	Περιλήφθηκαν αλλαγές 008/15-040\$a. Περιλήφθηκαν και αλλαγές για διαφορές μεταξύ 046 και 100\$a. Δεν καταμετρήθηκαν γιατί έγινε ο έλεγχος σε μία-μία εγγραφή. Διόρθωση/ενημέρωση 670\$a - (263 εγγραφές). Διαγραφή 670 με μόνη λέξη Τεκμήριο (419 εγγραφές). Δεν έγινε έλεγχος indicators
3/4/2023	19/4/2023	126	44	186	91	1662	165	2.274	51 παλαιές εγγραφές με LDR/6 a --> άλλαξε σε z 172 εγγραφές με διορθώσεις στο κείμενο του 663\$a 501 εγγραφές αλλαγή προσδιοριστή στο 663\$b 360 εγγραφές -- έλεγχος/προσθήκη 670\$a για βιβλioneet 62 εγγραφές -- διαγραφή υποπεδίου 670\$b με μόνο κείμενο ATEVE 865 εγγραφές -- διαγραφή υποπεδίου 0165z με τιμή A στα 3XX οι αλλαγές αφορούν σε προσθήκη τιμών για γλώσσες
27/4/2023	2/5/2023	185	68	474	267		284	1.278	Δείκτες -- διορθώθηκαν παλιές εγγραφές με λάθος δείκτη στο 400

Από τον Ιανουάριο του 2023 έως και το 1^ο δεκαπενθήμερο του Σεπτεμβρίου έχουν πραγματοποιηθεί συνολικά 33.302 διορθώσεις και εμπλουτισμοί στις εγγραφές Φυσικών προσώπων και Ψευδωνύμων (βασικό πεδίο 100), ενώ στις εγγραφές θεμάτων (βασικό πεδίο 150) έχουν γίνει περισσότερες από 227.000 διορθώσεις. Επιπλέον, έχει προστεθεί ο αντίστοιχος αγγλικός όρος από τις Θεματικές Επικεφαλίδες της Βιβλιοθήκης του Κογκρέσου (450\$7eng) σε περισσότερες από 28.000 εγγραφές, όπως επίσης ιεραρχικές σχέσεις (550\$x) σε περισσότερες από 8.000 εγγραφές και κλάσεις Dewey (083\$a) σε περισσότερες από 6.300 εγγραφές.

Συμπεράσματα

Η διασφάλιση ποιότητας στα βιβλιογραφικά δεδομένα είναι μία διαδικασία που περιλαμβάνει ποικίλα στάδια. Αρχικά, διαμορφώνεται η ροή εργασίας και παραμετροποιούνται τα συστήματα προκειμένου να αποφευχθεί η δημιουργία νέων ή επαναλαμβανόμενων λαθών. Στη συνέχεια, σε διοικητικό επίπεδο, αποφασίζεται η οπτική προσέγγιση και έπειτα τα κριτήρια ελέγχου. Ακολουθεί η επιλογή σημείων (πεδίων MARC21 εν προκειμένω) ελέγχου και η μέθοδος (π.χ. σύγκριση με ελεγχόμενες τιμές ενός λεξιλογίου ή με άλλο πεδίο). Έπειτα, αποφασίζεται ο τρόπος χειρισμού των λαθών ή των ευκαιριών εμπλουτισμού που παρουσιάζονται.

Η εφαρμογή των αλλαγών στη ροή εργασίας και η ανάπτυξη μεθοδολογίας συστηματικών διορθώσεων και εμπλουτισμών έχουν βελτιώσει την ποιότητα των εγγραφών και βοηθούν στην ανίχνευση περισσότερων ευκαιριών εμπλουτισμού. Για παράδειγμα, η προσθήκη των αγγλικών όρων στο 450\$7 των εγγραφών θεμάτων επέτρεψε την προσθήκη ιεραρχιών στο 550\$x και κλάσεων

Dewey στο 083ζα. Επιπλέον, διευκολύνουν τη μετατροπή των δεδομένων σε συνδεδεμένα δεδομένα και την ανάπτυξη καινοτόμων προβολών, όπως για παράδειγμα, την ενιαία προβολή της περιγραφής ενός συγγραφέα και όλων των ψευδωνύμων του.

Το Τμήμα Καταλόγων της ΕΒΕ συνεχίζει τη διαδικασία των ελέγχων και την εισαγωγή αυτόματων διαδικασιών, με στόχο τη μέγιστη δυνατή διευκόλυνση της εργασίας της καταλογογράφησης. Στα εργαλεία που εξετάζονται είναι εφαρμογές τεχνητής νοημοσύνης, όπως και το εργαλείο QA Catalogue που ανέπτυξε ο Péter Király (2023) για την ποιοτική ανάλυση εγγραφών. Η διαδικασία ελέγχου ποιότητας αναμένεται να εφαρμοστεί συστηματικά, μέσα στο 2025, στις βιβλιογραφικές εγγραφές (οντότητα Εκδήλωση) και να επεκταθεί στις εγγραφές συλλογικών οργάνων.

Βιβλιογραφία

- Bade, D. (2008). The Perfect Bibliographic Record: Platonic Ideal, Rhetorical Strategy or Nonsense? *Cataloging & Classification Quarterly*, 46(1), 109–133. <https://doi.org/10.1080/01639370802183081>
- Badovinac, B. (2017) Izhodišča za proučevanje kakovosti Podatkov V bibliografskih in normativnih zapisih: Kakovost Podatkov V Kontekstu in Raziskovalne USMERITVE V Katalogizaciji. *Knjižnica: revija za področje bibliotekarstva in informacijske znanosti*, 61(1–2). doi: [10.55741/knj.61.1-2.13847](https://doi.org/10.55741/knj.61.1-2.13847).
- Badovinac, B. (2021). Defining Data Quality in Bibliographic and Authority Records: A Case Study of the COBISS.SI System. *Cataloging & Classification Quarterly*, 59(4), 355–377. doi:[10.1080/01639374.2021.1910603](https://doi.org/10.1080/01639374.2021.1910603)
- Boissonnas, C. (1979). The quality of OCLC bibliographic records: the Cornell Law Library experience. *Law Library Journal* 72(1), 80-85.
- Bruce, T.R., & Hillmann, D.I. (2004). The Continuum of Metadata Quality: Defining, Expressing, Exploiting. Σε Hillmann, D.I., & Westbrook, E.L., ed. (2004). *Metadata in Practice* (238-256). Chicago: American Library Association, 2004. <https://hdl.handle.net/1813/7895>
- Gilliland, A. (2016). Setting the Stage. Σε Baca, M., ed. (2016). Introduction to metadata. 3rd ed. Los Angeles: Getty Publications. <http://www.getty.edu/publications/intrometadata>
- Hillmann, D.I. (2008) 'Metadata quality: From evaluation to augmentation', *Cataloging & Classification Quarterly*, 46(1), pp. 65–80. doi:[10.1080/01639370802183008](https://doi.org/10.1080/01639370802183008).
- Király, P. (2023). *QA Catalogue: A data quality dashboard for MARC catalogues*. Github. <https://github.com/pkiraly/qa-catalogue-web>
- Kumar, V., Chandrappa, & Harinarayana, N. (2024). Exploring dimensions of metadata quality assessment: A scoping review. *Journal of Librarianship and Information Science*. doi:[10.1177/09610006241239080](https://doi.org/10.1177/09610006241239080).
- Snow, K. (2017). Defining, Assessing, and Rethinking Quality Cataloging. *Cataloging & Classification Quarterly*, 55(7–8), 438–455. doi:[10.1080/01639374.2017.1350774](https://doi.org/10.1080/01639374.2017.1350774)