



Μελέτη για την ανάλυση και αξιολόγηση της ποιότητας των δεδομένων που παρέχονται από το Εθνικό Μητρώο Διοικητικών Διαδικασιών (ΕΜΔΔ)

έκδοση 1.0

Ιστορικό Εγγράφου

Ο παρακάτω πίνακας παρέχει μια επισκόπηση των πιο σημαντικών αλλαγών του εγγράφου.

Ημερομηνία	Έκδοση	Τροποποίηση	Συγγραφείς
31/07/2023	1.0	Αρχική έκδοση	Καπάνταη Ελένη, Μιχαηλίδης Αλέξιος, Πατρίκιος Γεώργιος, Μαγνήσαλης Ιωάννης Μπερμπερίδης Χρήστος Περιστεράς Βασίλειος Τεκτονίδης Απόστολος Τεντσογλίδης Ιορδάνης

Πίνακας Περιεχομένων

Ιστορικό Εγγράφου	2
Πίνακας Περιεχομένων.....	3
Κατάλογος Πινάκων, Διαγραμμάτων, Εικόνων & Γραφημάτων	5
Κατάλογος Πινάκων.....	5
Κατάλογος Εικόνων	5
Επιτελική Σύνοψη.....	6
1. Εισαγωγή.....	7
1.1. Ορισμός του προβλήματος	7
1.2. Σκοπός της μελέτης	9
1.3. Οργάνωση του κειμένου	10
2. Ανάλυση και αξιολόγηση της ποιότητας κυβερνητικών δεδομένων.....	10
2.1. Μεθοδολογία.....	11
2.2. Καλές πρακτικές στη διεθνή βιβλιογραφία.....	14
2.2.1. Πλαίσια.....	14
2.2.2. Κριτήρια και Μετρικές	15
2.2.3. Περιπτώσεις χρήσης.....	17
2.2.4. Αρχές Ανοιχτών Κυβερνητικών Δεδομένων (OGD Principles).....	20
2.2.5. Fair Principles.....	21
3. Εθνικό Μητρώο διαδικασιών.....	22
3.1. Ανάλυση δεδομένων.....	24
4. Ερευνητική Προσέγγιση	28
4.1. Δεδομένα που χρησιμοποιήθηκαν.....	29

4.2.	Μοντέλο ποιοτικής ανάλυσης.....	29
4.3.	Τεχνική Υλοποίηση και Αποτελέσματα.....	34
4.3.1.	Επιλογή των πεδίων στα οποία θα εφαρμοστούν οι προκαθορισμένες μετρικές.	35
4.3.2.	Δημιουργία δυναμικού dashboard για την ανάλυση και την οπτικοποίηση των αποτελεσμάτων μας.....	36
5.	Συμπεράσματα	47
5.1.	Σύντομη Επισκόπηση.....	47
5.2.	Κατεύθυνση επόμενων βημάτων	48
6.	Βιβλιογραφία	49
	Παράρτημα Α: Μετρικές Ανοιχτών Κυβερνητικών Δεδομένων.....	53
	Παράρτημα Β: Μονάδες Μέτρησης Ποιοτικών Μετρικών	54
	Παράρτημα Γ: Πεδία ανά μετρική.....	54

Κατάλογος Πινάκων, Διαγραμμάτων, Εικόνων & Γραφημάτων

Κατάλογος Πινάκων

Πίνακας 1: Μοντέλο ποιοτικής αξιολόγησης για τα δεδομένα του ΕΜΔΔ	32
Πίνακας 2: Πεδία του ΕΜΔΔ που αξιολογήθηκαν ανά μετρική	35

Κατάλογος Εικόνων

Εικόνα 1: Σχήμα απάντησης από το Mitos API για μια διαδικασία.....	23
Εικόνα 2: Σχήμα απάντησης δικαιολογητικών από το Mitos API για μια διαδικασία.....	24
Εικόνα 3: Ροή εργασίας.....	27
Εικόνα 4: Σελίδα πλοήγησης της δυναμικής αναφοράς	37
Εικόνα 5: Έλεγχος Ακρίβειας (Accuracy).....	38
Εικόνα 6: Λάθη Ακρίβειας που αφορούν διαδικασίες.....	39
Εικόνα 7: Έλεγχος Πληρότητας (Completeness) των υποχρεωτικών πεδίων.....	40
Εικόνα 8: Completeness των υποχρεωτικών πεδίων "description" & "org_owner"	41
Εικόνα 9: Πληρότητα (Completeness) όλων των πεδίων χωρίς ειδικές συνθήκες	42
Εικόνα 10: Έλεγχος πληρότητας των μη υποχρεωτικών πεδίων των διαδικασιών (αφορά τη μετρική Relevance)	43
Εικόνα 11: Έλεγχος εγκαιρότητας (Timeliness) βάσει του πεδίου "last_updated"	44
Εικόνα 12: Έλεγχος αξιοπιστίας (Credibility) βάσει του πεδίου «rule_url»	45

Εικόνα 13: Έλεγχος εγκυρότητας (Validity) των διευθύνσεων URL των διαδικασιών	45
Εικόνα 14: Έλεγχος Consistency μέσω Business Rules	46

Επιτελική Σύνοψη

Στα πλαίσια του ψηφιακού μετασχηματισμού της δημόσιας διοίκησης, η ποιότητα της πληροφορίας που παρέχεται από τις δημόσιες υπηρεσίες αποτελεί σημαντικό παράγοντα για την αποτελεσματική χρήση από τα ενδιαφερόμενα μέρη (π.χ πολίτες, επιχειρήσεις, δημόσιοι υπάλληλοι) και παράλληλα δείκτη επιτυχούς ηλεκτρονικής διακυβέρνησης. Μία από τις σημαντικότερες πηγές πληροφόρησης των πολιτών αποτελούν τα εθνικά μητρώα/αποθετήρια και περιλαμβάνουν δεδομένα που σχετίζονται με κυβερνητικές πολιτικές και τη λήψη αποφάσεων στον τομέα της ηλεκτρονικής διακυβέρνησης. Η διατήρηση της πληρότητας και της αξιοπιστίας της πληροφορίας που διατίθεται από τα εθνικά μητρώα αποτελεί βασικό μέλημα και πρόκληση για τις κυβερνήσεις και τους υφιστάμενους δημόσιους οργανισμούς.

Το παρόν κείμενο αποτελεί το πρώτο παραδοτέο του πακέτου εργασίας 1 και επικεντρώνεται στην ανάλυση και την αξιολόγηση των δεδομένων που παρέχονται από το Εθνικό Μητρώο Διοικητικών Διαδικασιών (ΕΜΔΔ) με σκοπό να εντοπιστούν και να καταγραφούν δυνατότητες και προτάσεις βελτίωσης της διαθέσιμης πληροφορίας. Ενδεικτικά η ανάλυση μπορεί να οδηγήσει στον εντοπισμό λαθών/ παραλείψεων στην καταχώρηση δεδομένων, συχνά επαναχρησιμοποιούμενων δεδομένων, καθώς επίσης και στην αναγνώριση συσχετίσεων μεταξύ των δεδομένων που μπορούν στη

συνέχεια να χρησιμοποιηθούν για την ομαδοποίηση ή την προτεραιοποίησή τους.

Το έργο ξεκινά με τη μελέτη των δεδομένων που παρέχονται από το ΕΜΔΔ, καθώς και τη μελέτη καλών πρακτικών στο αντικείμενο της αξιολόγησης της ποιότητας κυβερνητικών δεδομένων. Απώτερος στόχος είναι ο ορισμός και η εφαρμογή ενός πλαισίου ποιότητας στα δεδομένα του ΕΜΔΔ που θα ελέγχει την τρέχουσα κατάσταση των δεδομένων ως προς τις διαστάσεις του πλαισίου. Το τελικό παραδοτέο θα περιλαμβάνει έναν ολοκληρωμένο οδηγό ποιοτικής αξιολόγησης προσαρμοσμένο στις απαιτήσεις του ΕΜΔΔ, καθώς και ένα δυναμικό εργαλείο ανάλυσης και οπτικοποίησης της ποιότητας της πληροφορίας που εξετάζεται. Στη συνέχεια του έργου, παρουσιάζονται προτάσεις βελτίωσης σύμφωνα με τα αποτελέσματα της ανάλυσης.

1. Εισαγωγή

1.1. Ορισμός του προβλήματος

Τα τελευταία χρόνια παρατηρείται έντονο ενδιαφέρον για αναβάθμιση της Δημόσιας Διοίκησης με κύριες αναφορές στην απαλλαγή από τη γραφειοκρατία, καθώς και τις χρονοβόρες και πολύπλοκες διαδικασίες. Στο πλαίσιο αυτό οι κυβερνήσεις τείνουν να αναβαθμίσουν την παροχή υπηρεσιών στους πολίτες και τις επιχειρήσεις μέσω των συστημάτων ηλεκτρονικής διακυβέρνησης επιδιώκοντας την αποτελεσματική συνεργασία και την ανταλλαγή δεδομένων μεταξύ των συναρμόδιων φορέων. Σε αυτή την προσπάθεια καταλυτικός είναι ο ρόλος της ποιότητας της πληροφορίας (data quality) που παράγεται είτε από τα συστήματα ηλεκτρονικής διακυβέρνησης, είτε διατίθεται από τις σχετικές βάσεις δεδομένων (π.χ. μητρώα) καθώς ενισχύει την εμπιστοσύνη και ικανοποίηση του κοινού ως προς τις αντίστοιχες ψηφιακές υπηρεσίες, ενώ

παράλληλα δίνεται η δυνατότητα για ποιοτικότερη περαιτέρω επεξεργασία της πληροφορίας που παρέχεται.

Η διασφάλιση της ποιότητας δεδομένων είναι ένα πολυδιάστατο πρόβλημα το οποίο καλείται επίκαιρο τουλάχιστον τα τελευταία 50 χρόνια απασχολώντας τόσο τον ιδιωτικό όσο και το δημόσιο τομέα και παρά το γεγονός ότι υπάρχουν πολλές και διαφορετικές προσεγγίσεις που εξετάζουν δεν υπάρχει μία κοινή αποδεκτή λύση. Στη βιβλιογραφία υπάρχουν διαφορετικοί ορισμοί και κριτήρια κατηγοριοποίησης που σχετίζονται με την αξιολόγηση της ποιότητας δεδομένων, τα οποία μπορεί να συμφωνούν μεταξύ τους σε μικρότερο ή μεγαλύτερο βαθμό και εμφανίζουν μεγάλη συσχέτιση με το είδος της πληροφορίας στην οποία εφαρμόζονται καθιστώντας ανέφικτο τον ορισμό ενός κοινού δια-τομεακού πλαισίου ποιότητας. Έτσι για παράδειγμα σε επίπεδο δημόσιων κυβερνητικών δεδομένων το πρόβλημα είναι εντονότερο συγκριτικά με τον ιδιωτικό τομέα, καθώς τα κανονιστικά, νομοθετικά πλαίσια και οι διαδικασίες δεν μπορούν να μεταβληθούν με την ίδια ευκολία στο πλαίσιο βελτίωσης της πληροφορίας, δεδομένης και της αλληλεξάρτησης των δημόσιων οργανισμών.

Τα δεδομένα τα οποία παράγονται και διατίθενται από τους φορείς της δημόσιας διοίκησης αναφέρονται ως κυβερνητικά δεδομένα (Government Data). Τα τελευταία χρόνια το ενδιαφέρον έχει στραφεί στην έννοια των ανοιχτών κυβερνητικών δεδομένων (Open Government Data), τα οποία διανέμονται μέσω ηλεκτρονικών ιστοσελίδων και κυβερνητικών πυλών, με σκοπό τη δημιουργία χρήσιμης, εύκολα προσβάσιμης και επαναχρησιμοποιούμενης πληροφορίας, τόσο από υπολογιστικά συστήματα όσο και από τον απλό χρήστη. Προς αυτή την κατεύθυνση διάφορες πρωτοβουλίες και πλαίσια έχουν προκύψει με σκοπό τον έλεγχο και τη διασφάλιση της ποιότητας ανοιχτών κυβερνητικών δεδομένων, δίνοντας έμφαση κυρίως σε οργανωτικά και τεχνολογικά ζητήματα

που σχετίζονται με το εκάστοτε πληροφοριακό σύστημα (όπως για παράδειγμα το είδος των μεταδεδομένων που χαρακτηρίζουν ένα σύνολο δεδομένων) και αγνοώντας τα σφάλματα, τις ελλείψεις ή ανακρίβειες στις ίδιες τις τιμές των δεδομένων, γεγονός το οποίο δεν εξασφαλίζει την παραγωγή πληροφοριών υψηλής ποιότητας όσον αφορά το περιεχόμενό τους.

1.2. Σκοπός του παραδοτέου

Στο παρόν κείμενο ασχολούμαστε με την ανάλυση και αξιολόγηση της ποιότητας της πληροφορίας του Εθνικού Μητρώου Διοικητικών Διαδικασιών (ΕΜΔΔ) όπως αυτή διατίθεται από το API του φορέα.

Σκοπός του παραδοτέου είναι η ανάκτηση και η ανάλυση των δεδομένων που μπορούν να αξιολογηθούν με αυτοματοποιημένο τρόπο, λαμβάνοντας υπόψιν καλές πρακτικές και διαθέσιμες μετρικές που προτείνονται στη βιβλιογραφία.

Οι επιμέρους στόχοι του παραδοτέου είναι οι εξής:

- Μελέτη της βιβλιογραφίας και αναζήτηση καλών πρακτικών στον τομέα της ανάλυσης και αξιολόγησης της ποιότητας κυβερνητικών δεδομένων που παρέχονται αποκλειστικά για πληροφόρηση.
- Μελέτη του Εθνικού Μητρώου Διαδικασιών (ΕΜΔΔ) ως προς την οργάνωση και το περιεχόμενο της πληροφορίας που παρέχει
- Ανάλυση απαιτήσεων και παρουσίαση της μεθοδολογικής προσέγγισης που προτείνεται για την αξιολόγηση της ποιότητας των δεδομένων του ΕΜΔΔ
- Έλεγχος της ποιότητας των δεδομένων με χρήση εργαλείου ανάλυσης και οπτικοποίησης δεδομένων ως προς τα ποιοτικά χαρακτηριστικά που μπορούν να μετρηθούν υπολογιστικά

1.3. Οργάνωση του κειμένου

Στο Κεφάλαιο 2 εξετάζονται υφιστάμενα πλαίσια, μετρικές και διαστάσεις στο πλαίσιο αξιολόγησης ποιότητας κυβερνητικών δεδομένων

Στο Κεφάλαιο 3 παραθέτονται και αναλύονται πληροφορίες σχετικές με τη συλλογή, προεπεξεργασία και ανάλυση των δεδομένων του Εθνικού Μητρώου Διοικητικών Διαδικασιών (ΕΜΔΔ).

Στο Κεφάλαιο 4 παρουσιάζεται η μεθοδολογία που θα ακολουθηθεί για την αξιολόγηση των δεδομένων του ΕΜΔΔ σύμφωνα με τη δομή, τις απαιτήσεις και τις ανάγκες του ΕΜΔΔ και παρουσιάζεται η τεχνική υλοποίηση του προτεινόμενου μοντέλου ποιότητας, καθώς και τα αποτελέσματα που προκύπτουν.

Τέλος, στο Κεφάλαιο 5 ακολουθεί σύνοψη των κυριότερων ευρημάτων της παρούσας μελέτης και δίνονται οδηγίες για τα επόμενα βήματα.

2. Ανάλυση και αξιολόγηση της ποιότητας κυβερνητικών δεδομένων

Υπάρχουν πολλοί διαφορετικοί ορισμοί γύρω από την έννοια της ποιότητας δεδομένων «data quality» (Batini et al., 2009), αλλά κατά βάση αντιμετωπίζεται ως ο βαθμός στον οποίο τα δεδομένα που εξετάζονται ικανοποιούν συγκεκριμένες προϋποθέσεις (Lacagnina et al., 2023). Οι προϋποθέσεις αυτές μπορεί να διαφέρουν ανάλογα με τον τύπο των δεδομένων, τους ενδιαφερόμενους (stakeholders) ή τον σκοπό χρήσης. Υπό αυτό το πρίσμα είναι γνωστοί οι ορισμοί «fit-for-use» (τα δεδομένα είναι κατάλληλα για την προβλεπόμενη χρήση και απαλλαγμένα από λάθη και ανακρίβειες) ή «fit-for-purpose» (τα δεδομένα εξυπηρετούν τις ανάγκες και απαιτήσεις των χρηστών) (Lacagnina et al., 2023, Redman, 2001, Wang & Strong, 2015).

Η έννοια του «data quality» μπορεί να μεταφραστεί ποικιλοτρόπως: μπορεί να αφορά τη διαχείριση δεδομένων βάσει συγκεκριμένων προτύπων όπως ορίζεται για παράδειγμα από το πρότυπο ISO 8000, μπορεί να συνδέεται άμεσα με συγκεκριμένα κριτήρια (accuracy, timeliness, etc.) ή μπορεί να συνδέεται με την καταλληλότητα των δεδομένων βάσει ορισμένων κριτηρίων ως προς την (επανα-)χρησιμοποίησή τους για ένα συγκεκριμένο σκοπό. Οι διάφορες προσεγγίσεις και πρακτικές που προτείνονται στη βιβλιογραφία αναλύονται στις ενότητες που ακολουθούν.

2.1. Μεθοδολογία

Η μελέτη της βιβλιογραφίας προσανατολίστηκε σε υλικό που επικεντρώνεται στις μεθόδους αξιολόγησης της ποιότητας της πληροφορίας που παρέχεται από δημόσιες υπηρεσίες. Για το σκοπό χρησιμοποιήθηκαν οι μηχανές αναζήτησης "Scopus", "WebOfScience" και "Google Scholar" εφαρμόζοντας τους ακόλουθους όρους: "data quality"/ "metadata quality"/ "data curation" σε συνδυασμό με "open government data" /"public data"/ "public government data"/ "government data"/ "e-government"/"public services".

Δεδομένου ότι το «data quality» είναι ένα πεδίο με μεγάλη απήχηση όχι μόνο στην ακαδημαϊκή κοινότητα, αλλά και στον ιδιωτικό και δημόσιο τομέα (δημόσιες αρχές, κυβερνητικοί οργανισμοί κλπ) η έρευνα επεκτάθηκε και σε πηγές εκτός των βασικών επιστημονικών βιβλιοθηκών συμπεριλαμβάνοντας αναφορές που δημοσιεύονται από διεθνείς οργανισμούς και ερευνητικά ιδρύματα (π.χ. Ευρωπαϊκή Επιτροπή, Διεθνής Κοινότητα Διαχείρισης Δεδομένων (DAMA International)).

Τα αποτελέσματα που προέκυψαν από την αναζήτηση καλύπτουν ένα αρκετά μεγάλο εύρος ερευνητικών μελετών μεταξύ των οποίων παρατηρήθηκε μεγάλη διαφοροποίηση στις μεθοδολογικές τους προσεγγίσεις, επιβεβαιώνοντας την πολυπλοκότητα του πεδίου. Οι πιο σημαντικές κατηγορίες υφιστάμενων

μελετών που ξεχώρισαν αφορούν: α) γενική έρευνα σχετικά με την ποιότητα των δεδομένων κυρίως προσανατολισμένες στον ορισμό ποιοτικών μετρικών και την ομαδοποίησή τους π.χ. (Sidi et al., 2012) β) πλαίσια ποιοτικής αξιολόγησης δεδομένων γ) αξιολόγηση ανοιχτών (κυβερνητικών) δεδομένων δ) αξιολόγηση διασυνδεδεμένων δεδομένων ε) οδηγούς για δημοσίευση ποιοτικών δεδομένων (EU Guidelines, 2021; EU Open Data Goldbook, 2018). Ως προς τη δεύτερη κατηγορία οι υφιστάμενες μελέτες μπορούν να κατηγοριοποιηθούν σε διαφορετικά επίπεδα αξιολόγησης ανάλογα με το είδος της πληροφορίας που ελέγχεται (υπηρεσίας «service quality», συνόλου δεδομένων «dataset quality», περιεχομένου «content quality») με επικρατέστερη την αξιολόγηση σε επίπεδο συνόλου δεδομένων (Andrade et al., 2021; Bouchelouche et al., 2022; Ijab et al., 2017; Kučera et al., 2013; Lämmel et al., 2020; Moreno et al., 2022; Nogueras-Iso et al., 2021; Raca et al., 2022a; Sanabria et al., 2018a). Στην περίπτωση αυτή αξιολογείται ένα σύνολο δεδομένων ως δημοσιευμένη εγγραφή σε κάποιο portal βάσει των εξωτερικών χαρακτηριστικών του (π.χ. τίτλος, εκδότης, ημερομηνία δημοσίευσης)¹. Σε επίπεδο content αναλύεται αποκλειστικά η περιεχόμενη πληροφορία ενός συνόλου δεδομένων. Αναφορικά με το επίπεδο service εξετάζονται system-based ή webdesign χαρακτηριστικά όπως για παράδειγμα ο βαθμός απόκρισης σε κάποια εντολή του χρήστη, χαρακτηριστικά διεπαφής χρήστη, κτλ, και σε μικρότερο βαθμό η ποιότητα της παρεχόμενης πληροφορίας.

Είναι σημαντικό να σημειωθεί ότι ορισμένες μελέτες μπορεί να εμπίπτουν σε περισσότερες από μία κατηγορίες. Για παράδειγμα, οι αναφορές που σχετίζονται με την αξιολόγηση σε επίπεδο συνόλου δεδομένων, στην πλειοψηφία τους αφορούν ανοιχτά κυβερνητικά δεδομένα.

¹ Στη βιβλιογραφία το επίπεδο αυτό αναφέρεται συχνά ως portal-level data quality

Η έννοια της αξιολόγησης της ποιότητας δεδομένων μπορεί επίσης να εντάσσεται και στο ευρύτερο πλαίσιο της διαχείρισης και της οργάνωσης των δεδομένων ενός οργανισμού δίνοντας έμφαση στις προϋποθέσεις και τους παράγοντες που καθορίζουν την παραγωγή ποιοτικών συνόλων δεδομένων, (Ijab et al., 2017). Μία σημαντική πρόκληση που αντιμετωπίστηκε στο στάδιο μελέτης της βιβλιογραφίας αφορά το γεγονός ότι η έννοια “data quality” συχνά χρησιμοποιείται ως ισοδύναμη με την έννοια «metadata quality». Το ίδιο πρόβλημα παρατηρήθηκε και με τις έννοιες «data quality», «data quality management» και «data quality governance», δυσκολεύοντας σε ορισμένες περιπτώσεις τη διαδικασία εντοπισμού και επιλογής κατάλληλου υλικού προς αξιολόγηση.

Προκειμένου να μειωθεί το εύρος της πληροφορίας που συλλέχθηκε και να ευθυγραμμιστεί με το στόχο του συγκεκριμένα έργου, επιλέχθηκαν μελέτες σύμφωνα με τα εξής κριτήρια:

- Μελέτες που περιλαμβάνουν πληροφορία προσανατολισμένη στη φάση αξιολόγησης/ ελέγχου (assessment/monitoring aspects) όπου διερευνήθηκαν καλές πρακτικές (μέθοδοι, μετρικές και πρότυπα) που χρησιμοποιούνται στο πλαίσιο αυτό.
- Μελέτες που αξιολογούν την ποιότητα της παρεχόμενης πληροφορίας σε επίπεδο περιεχομένου (content). Στην περίπτωσή μας ως περιεχόμενο ορίζεται η πληροφορία του MITOS που σχετίζεται με τις περιγραφές των διαδικασιών (πεδία (attributes), τιμές πεδίων (data values)), τα οποία εφεξής θα αναφέρονται ως metadata και data αντίστοιχα .
- Μελέτες που εξετάζουν την ποιότητα διασυνδεδεμένων δεδομένων (linked data) δεν αποτελούν αντικείμενο ενδιαφέροντος του συγκεκριμένου παραδοτέου.

2.2. Καλές πρακτικές στη διεθνή βιβλιογραφία

2.2.1. Πλαίσια

Η σημαντικότητα της δημιουργίας και διατήρησης δεδομένων υψηλής ποιότητας έχει οδηγήσει στην ανάπτυξη ποικίλων μεθοδολογιών και πλαισίων που αφορούν τη διαχείριση της ποιότητας δεδομένων και στοχεύουν στην αξιολόγηση και βελτίωση της ποιότητας των δεδομένων. Οι (Batini et al., 2009) παρουσιάζουν και συγκρίνουν μερικές από τις δημοφιλέστερες προσεγγίσεις προς αυτή την κατεύθυνση αξιολογώντας διάφορα κριτήρια όπως για παράδειγμα το είδος των δεδομένων και των πληροφοριακών συστημάτων, κόστος και ποιοτικές μετρικές. Συμπεραίνουν ότι πολλές μεθοδολογίες που προτείνονται στη βιβλιογραφία έχουν αναπτυχθεί βασιζόμενες στις ιδιαιτερότητες ενός τομέα (domain-specific), γεγονός που περιορίζει την εφαρμογή τους. Η επιλογή των διαστάσεων και των τεχνικών Data Quality (DQ) που αφορούν την αξιολόγηση της ποιότητας των δεδομένων είναι κατά κανόνα προσαρμοσμένη στο εκάστοτε σενάριο και, ως εκ τούτου, δεν μπορεί να υπάρξει μία θεωρία με γενική εφαρμογή (Nikiforova, 2020).

Στο ίδιο πλαίσιο οι (Cichy & Rass, 2019) εξετάζουν τις υφιστάμενες θεωρίες σε τρία επίπεδα (ορισμός, αξιολόγηση, βελτίωση) προτείνοντας έναν οδηγό απόφασης ως προς την επιλογή του κατάλληλου πλαισίου ποιότητας ανάλογα με τις απαιτήσεις της κάθε περίπτωσης. Μία από τις σημαντικότερες θεωρίες του πεδίου που αξίζει να αναφερθεί είναι η TDQM², καθώς θεωρείται ότι μπορεί να εφαρμοστεί σε διάφορα context και είναι η μοναδική περίπτωση που λαμβάνει υπόψιν τους επιχειρησιακούς κανόνες (business rules) ενός οργανισμού για την αξιολόγηση της πληροφορίας του. Ενδιαφέρον παρουσιάζει και το πλαίσιο DQAF³ (Sebastian-Coleman, 2013) που περιγράφει τρία διαφορετικά επίπεδα αξιολόγησης: αρχική αξιολόγηση (Initial Assessment), έλεγχοι διαδικασιών (Process Controls), περιοδική μέτρηση (Periodic

² Total Quality Data Management

³ Data Quality Assessment Framework

Measurement). Παρά το ότι κάθε επίπεδο έχει διαφορετικό στόχο, ακολουθείται σε όλα κοινή αλληλουχία βημάτων για την αξιολόγηση των δεδομένων.

Υπενθυμίζεται ότι τα πλαίσια που παρουσιάζονται στις εν λόγω μελέτες εξετάστηκαν μόνο ως προς τον τρόπο που προσεγγίζουν το στάδιο ανάλυσης και αξιολόγησης και παραθέτονται για μελλοντική διερεύνηση ως προς την ανάπτυξη ενός γενικότερου ποιοτικού πλαισίου διαχείρισης ποιότητας των δεδομένων του ΕΜΔΔ.

2.2.2. Κριτήρια και Μετρικές

Η πιο διαδεδομένη προσέγγιση αξιολόγησης της ποιότητας των δεδομένων είναι η εφαρμογή διαστάσεων/μετρικών ποιότητας. Οι μετρικές αυτές εξετάζονται θεσπίζοντας κατάλληλες μονάδες μέτρησης και μπορεί να είναι είτε αντικειμενικές (π.χ. συμμόρφωση με συγκεκριμένους κανόνες), είτε υποκειμενικές (π.χ. βαθμός ικανοποίησης χρηστών). Στο παρόν έργο μελετώνται αποκλειστικά αντικειμενικές μετρικές αξιολόγησης.

Μία από τις βασικότερες προκλήσεις του πεδίου και κατ' επέκταση και της παρούσας μελέτης αποτελεί η ύπαρξη πολλών και διαφορετικών, συχνά αλληλεπικαλυπτόμενων μετρικών (μετρικές με ίδια ονομασία και διαφορετική επεξήγηση, αλλά και το αντίστροφο). Ανομοιογένεια παρατηρήθηκε επίσης και στις κατηγοριοποιήσεις των μετρικών. Ενδεικτικά, η πιο δημοφιλής ταξινόμηση προτάθηκε από τους (Wang & Strong, 2015) οι οποίοι κατηγοριοποίησαν 15 ποιοτικές μετρικές στις κατηγορίες: εγγενώς (intrinsic), σύμφωνα με το πλαίσιο εφαρμογής (contextual), αναπαραστατικά (representational), προσβασιμότητα (accessibility). Οι (Redman, 2001) προτείνουν μία πιο ευέλικτη και κατανοητή ταξινόμηση, χρησιμοποιώντας 3 κριτήρια ομαδοποίησης: εννοιολογικό σχήμα (conceptual schema), τιμές δεδομένων (data values), δομή δεδομένων (data format). Σύμφωνα με τους (Wand & Wang, 1996) οι πιο σημαντικοί δείκτες ποιότητας δεδομένων είναι οι: πληρότητα (completeness), ακρίβεια (accuracy), αξιοπιστία (reliability), συνέπεια (consistency) και εγκαίριότητα (timeliness). Οι μετρικές πληρότητα (completeness), ακρίβεια (accuracy), εγκαίριότητα

(timeliness) χρησιμοποιούνται σχεδόν από όλα τα μοντέλα αξιολόγησης (Vetrò et al., 2016).⁴

Ένας από τους πιο καινούριους ορισμούς ποιοτικών μετρικών προτάθηκε από τη διεθνή κοινότητα DAMA⁵ το 2013 και περιλαμβάνει 6 ποιοτικές μετρικές (completeness, uniqueness, timeliness, validity, accuracy, consistency) αποτελώντας οδηγό για πολλές ερευνητικές μελέτες, αλλά και για τον καθορισμό πλαισίου ποιότητας κυβερνητικών δεδομένων, όπως για παράδειγμα στην περίπτωση της Βρετανίας (Diamond & Chisholm, 2020). Η ολλανδική ομάδα εργασίας DAMA NL⁶ σε πιο πρόσφατες μελέτες επικεντρώνεται στην ομογενοποίηση των ποιοτικών μετρικών ως προς τον ορισμό τους λαμβάνοντας υπόψιν το είδος των δεδομένων που εξετάζονται καθώς και σημαντικές αναφορές (π.χ. ISO standards) και παρέχει σημαντικούς οδηγούς και λεξικά.

Τα ISO standards αποτελούν μία άλλη σημαντική πηγή πληροφόρησης για την αξιολόγηση της ποιότητας των δεδομένων καλύπτοντας διαφορετικές οπτικές του πεδίου. Αντιπροσωπευτικά παραδείγματα αποτελούν τα ISO 8000 και ISO 2500. Το ISO 8000- είναι το διεθνές πρότυπο για την ποιότητα των δεδομένων και περιλαμβάνει διαστάσεις σχετικές με τη διαχείριση της ποιότητας των δεδομένων. Η σειρά ISO 2500 σχετίζεται με την αξιολόγηση της ποιότητας δεδομένων πληροφοριακών συστημάτων, προτείνοντας μετρικές και διαστάσεις σε 2 επίπεδα: εγγενής ποιότητα δεδομένων (inherent data quality) και βασιζόμενη σε χαρακτηριστικά του συστήματος (system-dependent data quality). Το πρώτο επίπεδο αφορά το βαθμό στον οποίο τα δεδομένα καλύπτουν τις ανάγκες του πλαισίου στο οποίο εφαρμόζονται και περιλαμβάνει τις εξής μετρικές: ακρίβεια (accuracy), πληρότητα (completeness), συνέπεια (consistency), αξιοπιστία (credibility), εγκαιρότητα (up-to-dateness) (datos.gob.es, 2022).

⁴ Για τους σκοπούς του συγκεκριμένου έργου περιοριζόμαστε σε απλή αναφορά των διαστάσεων/μετρικών. Η παράθεση και ανάλυση των ορισμών ξεπερνάει το αντικείμενο μελέτης και τα χρονικά πλαίσια του συγκεκριμένου παραδοτέου

⁵ Data Management Association International UK Working Group

⁶ [Data quality dimension \[Data Management Wiki\]](#)

Σε κάθε περίπτωση, η έλλειψη μίας κοινής πρότυπης λίστας διαστάσεων και αντίστοιχων μονάδων μέτρησης δημιουργεί ένα πολύπλοκο περιβάλλον έρευνας το οποίο αποδίδεται στις διαφορετικές απαιτήσεις του κάθε οργανισμού και τομέα, και καθιστά σχεδόν αδύνατη τη δημιουργία ολιστικών λύσεων.

Οι (Nikiforova et al., 2020) λαμβάνοντας υπόψη τη σχετικότητα και την πολυπλοκότητα των ποιοτικών μετρικών ως μέθοδο αξιολόγησης ποιότητας προτείνουν τη χρήση πιο γενικών διαστάσεων και συγκεκριμένα τον καθορισμό «data quality requirements». Στο **Error! Reference source not found.** παρουσιάζονται οι πιο κοινές προϋποθέσεις σύμφωνα με τη μελέτη τους.

2.2.3. Περιπτώσεις χρήσης

Όπως ήδη έχει αναφερθεί η αξιολόγηση των ποιότητας των δεδομένων μπορεί να υφίσταται σε διαφορετικά επίπεδα π.χ. υπηρεσίας, συνόλου δεδομένων, περιεχόμενου. Για παράδειγμα, σε επίπεδο περιεχομένου, οι (Torchiano et al., 2017) αξιολογούν την ποιότητα δημόσιων συμφωνητικών και συμβάσεων που δημοσιεύτηκαν στις επίσημες ιστοσελίδες 12 Ιταλικών Πανεπιστημίων, με οδηγό τα ποιοτικά χαρακτηριστικά που ορίζονται από το ISO 25012: completeness (αφορά ελλείψεις τιμών ανά πεδίο ή εγγραφή), accuracy (ο βαθμός στον οποίο τα δεδομένα συμμορφώνονται με αναμενόμενες ή προκαθορισμένες μορφές), consistency (μελετά τις ασυνέπειες των δεδομένων σε επίπεδο οντότητας αλλά και μεταξύ δεδομένων που ανήκουν στο ίδιο πεδίο, αλλά διαφορετικό dataset). Οι ασυνέπειες σε επίπεδο οντότητας μπορεί να αφορούν λογικούς περιορισμούς (κανόνες) που έχουν οριστεί στα δεδομένα μετά από προσεκτική μελέτη του τομέα. Στο ίδιο πλαίσιο οι (Sanabria et al., 2018b) χρησιμοποίησαν ως δείκτες ποιότητας τις μετρικές completeness, accuracy, duplicity σε διαθέσιμα κοινοτικά δεδομένα. Τέλος, οι (Soylu et al., 2022) παρουσιάζουν μία διαφορετική προσέγγιση εξάγοντας και αναλύοντας ευρωπαϊκά δεδομένα σχετικά με δημόσιες προμήθειες που είναι διαθέσιμα στη

μεγαλύτερη διεθνή βάση στον τομέα αυτό⁷. Ο έλεγχος ποιότητας πραγματοποιείται σε επίπεδο περιεχομένου λαμβάνοντας υπόψιν ποιοτικές αστοχίες που μπορεί να προκύψουν (π.χ. ελλείψεις τιμών, διπλότυπες εγγραφές, συντακτικά λάθη) χωρίς όμως αυτές να αντιστοιχίζονται σε συγκεκριμένες μετρικές.

Μια ακόμη προσέγγιση που αξίζει να αναφερθεί είναι το μοντέλο αξιολόγησης που προτείνεται στο γενικότερο πλαίσιο της DQMP⁸ μεθοδολογίας η οποία προσανατολίζεται στη διαχείριση δεδομένων δημοσίου της Εσθονίας. Οι (Terandi et al., 2017) αξιολογούν τις ποιοτικές μετρικές που προτείνονται από διαφορετικές πηγές και καταλήγουν ότι αυτές δεν καλύπτουν όλα τα χαρακτηριστικά που είναι σημαντικά για τα δεδομένα δημοσίου. Για παράδειγμα, η αρχή Once-Only παρά τη σημαντικότητά της στο δημόσιο τομέα δεν συγκαταλέγεται στα ποιοτικά χαρακτηριστικά των δεδομένων. Επιπλέον, αναφορικά με το πρότυπο ISO/IEC 25012 τονίζεται ότι δε λαμβάνουν υπόψιν μετρικές που σχετίζονται με χαρακτηριστικά σε επίπεδο πληροφοριακού συστήματος π.χ recoverability, portability. Τελικά καταλήγουν στις ακόλουθες μετρικές: accuracy, completeness, consistency, credibility, up-to-dateness, compliance, confidentiality, once-only principle, and non-redundancy, ακολουθώντας το πρότυπο ISO/IEC 25012 για τον ορισμό τους, όπου αυτό είναι εφικτό. Προκειμένου να βελτιώσουν τη σαφήνεια των ορισμών εξετάζουν ακόμη μία σημαντική διάσταση που αφορά τον βαθμό αυτοτέλειας κάθε μετρικής, καταγράφοντας για κάθε μία το βαθμό εξάρτησής της από τις υπόλοιπες μετρικές που συνθέτουν το προτεινόμενο πλαίσιο.

Όπως έχει ήδη αναφερθεί αναφορές που σχετίζονται με την αξιολόγηση σε επίπεδο περιεχομένου ενδέχεται να εντοπιστούν και στα άλλα επίπεδα. Για

⁷ <https://openopps.com/>

⁸ Data Quality Methodology for Public organizations

παράδειγμα, οι (Vetrò et al., 2016) προτείνουν ένα metric-based μοντέλο αξιολόγησης βασιζόμενοι σε ποιοτικά χαρακτηριστικά της μεθοδολογίας SPDQM⁹ (Moraga et al., 2009) και πραγματοποιούν ανάλυση Ιταλικών κυβερνητικών δεδομένων σε δύο επίπεδα: συνόλου δεδομένων και περιεχομένου, ορίζοντας για κάθε μετρική και τη μονάδα μέτρησης που εφαρμόζουν για την αξιολόγησή της. Σημειώνεται ότι η το πλαίσιο SPDQM ενσωματώνει χαρακτηριστικά του ISO 25012 και τα επεκτείνει.

Οι (Raca et al., 2022b) ασχολούνται με την ποιότητα των portals ανοιχτών κυβερνητικών δεδομένων, παρουσιάζοντας ένα πλαίσιο αξιολόγησης για τα καταγεγραμμένα σύνολα δεδομένων που αφορούν χώρες των δυτικών Βαλκανίων. Το πλαίσιο περιλαμβάνει 3 διαστάσεις: ανοιχτότητα (openness), σύνολο δεδομένων (dataset), δεδομένα (data). Η διάσταση data αφορά το περιεχόμενο του συνόλου δεδομένων και ελέγχεται ως προς:

1. την ύπαρξη τιμών για τα πεδία που περιλαμβάνει το σύνολο (completeness)
2. Τον εντοπισμό διπλότυπων εγγραφών (uniqueness)
3. Τη συνέπεια ως προς την αναμενόμενη μορφή των δεδομένων (consistency)
4. Τη συμμόρφωση των δεδομένων ως προς τη μορφή που ορίζεται από προκαθορισμένα standards (validity).

Σε κάθε μετρική περιγράφεται και η αντίστοιχη υπολογιστική συνάρτηση για τη μέτρησή της. Μία παρατήρηση που αξίζει να αναφερθεί είναι ότι για τον υπολογισμό της πληρότητας των δεδομένων λαμβάνονται υπόψιν όλα τα

⁹ Square-Aligned Portal Data Quality Model

συμπεριλαμβανόμενα πεδία που περιγράφουν μία εγγραφή, αποδίδοντας έμμεσα σε όλα τα πεδία ίδιο βαθμό σημαντικότητας.

Μία γενικότερη παρατήρηση αφορά την ανομοιογένεια στις ονομασίες των μετρικών επιβεβαιώνοντας τον αρχικό μας ισχυρισμό. Πιο συγκεκριμένα η μετρική consistency ορίζεται με τον ίδιο τρόπο με τον οποίο οι (Torchiano et al., 2017) ορίζουν την ποιοτική μετρική accuracy.

Τέλος, δεδομένου ότι τα μητρώα¹⁰ μπορούν να ενταχθούν σε ξεχωριστή κατηγορία δημόσιων υπηρεσιών ως υπηρεσίες παροχής πληροφοριών (σύμφωνα με την κατηγοριοποίηση των Jansen & Ølnes, 2016), συνυπολογίστηκαν στη μελέτη μας και μελέτες που σχετίζονται με την ποιότητα δημόσιων υπηρεσιών (Alexoroulos et al., 2023), δίνοντας έμφαση στον τρόπο που αντιμετωπίζουν την ποιότητα της πληροφορίας σε επίπεδο περιεχομένου και όχι σε χαρακτηριστικά που σχετίζονται με την ποιότητα εκτελέσιμων δεδομένων ή webdesign (Papadomichelaki & Mentzas, 2012). Η πληροφορία που περιλαμβάνεται στις μελέτες αυτού του επιπέδου συνήθως περιορίζεται σε απλή αναφορά ποιοτικών μετρικών χωρίς να παρέχονται περαιτέρω διευκρινίσεις ως προς τον ορισμό ή τη μέτρησή τους.

2.2.4. Αρχές Ανοιχτών Κυβερνητικών Δεδομένων (OGD Principles)

Όπως έχει ήδη αναφερθεί τα τελευταία χρόνια το ενδιαφέρον έχει στραφεί στην έννοια των ανοιχτών κυβερνητικών δεδομένων (Open Government Data). Λαμβάνοντας υπόψιν ότι το ΕΜΔΔ αποτελεί μία εν δυνάμει πηγή ανοιχτής πληροφορίας, επιλέξαμε να συμπεριλάβουμε και τις υφιστάμενες προσεγγίσεις (current state-of-the-art) που αφορούν ανάλυση ποιότητας ανοιχτών δεδομένων. Ένα από τα πιο γνωστά σχήματα ποιότητας ανοιχτών δεδομένων

¹⁰ Στην Ελλάδα τέτοιες πηγές πληροφόρησης μπορούν να θεωρηθούν οι πλατφόρμες ΜΙΤΟΣ, ΔΙΑΥΓΕΙΑ, ΤΟ ΕΘΝΙΚΟ ΤΥΠΟΓΡΑΦΙΟ, Η ΑΠΟΓΡΑΦΗ κλπ

είναι το «5 stars open data»,¹¹ και αποτελείται από πέντε επίπεδα ποιότητας που αναπαρίστανται ως αστέρια. Το σχήμα αυτό επικεντρώνεται αποκλειστικά στη μορφή που χρησιμοποιείται για τη δημοσίευση των δεδομένων, συνεπώς δεν μπορεί να χρησιμοποιηθεί για την αξιολόγηση της συνολικής ποιότητας ενός συνόλου δεδομένων. Το 2007, η ομάδα «The Open Government Working Group» δημιούργησε ένα πιο ολοκληρωμένο σύνολο 8 αρχών το οποίο επεκτάθηκε με μερικά ακόμη χαρακτηριστικά από την ομάδα Sunlight Foundation¹² θέτοντας τη βάση για την ανάπτυξη μιας διαδικασίας αξιολόγησης της ποιότητας των ανοικτών δεδομένων. Τέλος, το Open Data Charter¹³ αποτελεί μια συνεργασία μεταξύ κυβερνήσεων και οργανισμών που εργάζονται για να καταστήσουν τα δεδομένα ανοικτά και ελεύθερα διαθέσιμα προτείνοντας ένα πλαίσιο 6 αρχών. Τα πλαίσια και οι μετρικές τους είναι διαθέσιμα στο [Παράρτημα Α: Μετρικές Ανοικτών Κυβερνητικών Δεδομένων](#)

Ένας βασικός περιορισμός των αρχών αυτών είναι ότι εφαρμόζονται σε επίπεδο συνόλου δεδομένων θεωρώντας εξορισμού την ποιότητα του περιεχομένου υψηλή (Nikiforova, 2020).

2.2.5. Fair Principles

Οι αρχές της FAIR (FAIR principles) δημοσιεύτηκαν επίσημα για πρώτη φορά το 2016 (Wilkinson, et al., 2016) δίνοντας έμφαση στην εύκολη πρόσβαση και επαναχρησιμοποίηση των δεδομένων και μπορούν να εφαρμοστούν τόσο σε επίπεδο δεδομένων όσο και σε μεταδεδομένα. Σε αντίθεση με τις υφιστάμενες αρχές για τα ανοιχτά κυβερνητικά δεδομένα πλεονεκτούν στο ότι δεν προϋποθέτουν τα δεδομένα να είναι ανοικτά ή να παρέχονται δωρεάν. Λειτουργούν κυρίως ως οδηγός ως προς τον τρόπο που πρέπει να

¹¹ T. Berners-Lee, "Linked data-design issues", Tech. Rep., 2006

¹² <https://sunlightfoundation.com/wp-content/uploads/sites/2/2016/11/Ten-Principles-for-Opening-Up-Government-Data.pdf>

¹³ [Open Data: Learn more about the international ODC principles \(opendatacharter.net\)](#)

οργανώνονται τα δεδομένα ώστε να είναι εύκολα προσβάσιμα, κατανοητά και επαναχρησιμοποιήσιμα. Σημαντικοί χρηματοδοτικοί φορείς χρησιμοποιούν και προωθούν τα FAIR principles, πχ. Η Ευρωπαϊκή Επιτροπή αναφέρει τα FAIR Principles στη στρατηγική της “European Data Strategy (2020)” ως έναν τρόπο υλοποίησης της διαλειτουργικότητας. Στην πλατφόρμα data.europa.eu (DEU)¹⁴ υπάρχουν διαθέσιμοι οδηγοί για την δημιουργία και δημοσίευση δεδομένων υψηλής ποιότητας ακολουθώντας τις αρχές FAIR.

3. Εθνικό Μητρώο διαδικασιών

Το Εθνικό Μητρώο Διοικητικών Διαδικασιών (ΕΜΔΔ) «Μίτος» είναι το επίσημο μητρώο όλων των διαδικασιών του ελληνικού δημοσίου. Το μητρώο απευθύνεται σε ιδιώτες, επαγγελματίες και στελέχη της δημόσιας διοίκησης, που ενδιαφέρονται να πληροφορηθούν για τον τρόπο, χρόνο και κόστος υλοποίησης των διαδικασιών του δημόσιου τομέα. Υπάρχουν 4 τύποι διαδικασιών που μπορούν να αναζητηθούν από το χρήστη: δημοσιευμένες, προς δημοσίευση, προς έγκριση, υπό επεξεργασία.

Ένα βασικό πλεονέκτημα του ΕΜΔΔ είναι η παροχή API που επιτρέπει την εύκολη πρόσβαση και επεξεργασία του φάσματος των διαδικασιών. Το API (Application Programming Interface ή Διεπαφή Προγραμματισμού Εφαρμογών) Mitos προσφέρει την δυνατότητα άντλησης δεδομένων σχετικά για τις διαδικασίες όπως αυτές περιγράφονται στο Εθνικό Μητρώο Διοικητικών Διαδικασιών. Το Mitos API ακολουθεί την αρχιτεκτονική REST (Representational State Transfer) σύμφωνα με το openAPI 3.0. Η χρήση του Mitos API είναι γίνεται με την χρήση αιτημάτων HTTP (Hypertext Transfer Protocol).

¹⁴ [Data quality - Documentation of data.europa.eu \(DEU\) \(dataeuropa.gitlab.io\)](https://data.europa.eu)

Τα δεδομένα των διαδικασιών επιστρέφονται από το Mitos API με μια συγκεκριμένη δομή που ονομάζεται σχήμα. Το σχήμα έχει την ακόλουθη δομή όπως φαίνεται στην παρακάτω εικόνα.

```
Schemas

process {
  process > [...]
  process_conditions > [...]
  process_evidences > [...]
  process_evidences_cost > [...]
  process_provision_digital_locations > [...]
  process_rules > [...]
  process_steps > [...]
  process_steps_digital > [...]
  process_useful_links > [...]
}
```

Εικόνα 1: Σχήμα απάντησης από το Mitos API για μια διαδικασία

Το σχήμα αναπτύσσεται επιπλέον για όλα τα μέρη που αποτυπώνονται στην παραπάνω εικόνα όπως `process`, `process_evidences`, `process_evidences_cost` κλπ. Η Εικόνα 2 παρουσιάζει την ανάπτυξη για τα `process_evidences` όπως αυτά δομούνται σύμφωνα με το Mitos API.

```

Schemas

process {
  process
  process_conditions > [...]
  process_evidences
  process_evidences {
    process_evidence_alternative string
    process_evidence_alternative_of string
    process_evidence_description string
    process_evidence_is_under_prerequisite string
    process_evidence_note string
    process_evidence_num_id string
    process_evidence_owner string
    Enum:
    > Array [ 6 ]
    process_evidence_prerequisite string
    Enum:
    > Array [ 36 ]
    process_evidence_related_process string
    process_evidence_related_url string
    process_evidence_submission_type string
    Enum:
    > Array [ 7 ]
    process_evidence_type string
    Enum:
    > Array [ 90 ]
  }
  process_evidences_cost > [...]
  process_provision_digital_locations > [...]
  process_rules > [...]
  process_steps > [...]
  process_steps_digital > [...]
  process_useful_links > [...]
}

```

Εικόνα 2: Σχήμα απάντησης δικαιολογητικών από το Mitos API για μια διαδικασία

Στην παρούσα φάση η πρόσβαση είναι περιορισμένη και παρέχεται κατ' αποκλειστικότητα από τους διαχειριστές του MITOS δεδομένου ότι η πλατφόρμα βρίσκεται σε πιλοτικό στάδιο. Για τις ανάγκες του έργου παραχωρήθηκε πρόσβαση στο Mitos API για την άντληση των δεδομένων των διαδικασιών τα οποία αναλύθηκαν με την χρήση εργαλείων εξόρυξης και ανάλυσης δεδομένων.

3.1. Ανάλυση δεδομένων

Στην ενότητα αυτή περιγράφονται τα εργαλεία και οι βασικές ενέργειες που απαιτούνται για την ανάλυση των δεδομένων του ΕΜΔΔ, από τη συλλογή έως την οπτικοποίησή τους. Για την εξόρυξη και οπτικοποίηση των δεδομένων έγινε χρήση των ακόλουθων εργαλείων:

KNIME¹⁵

Το KNIME είναι μια πλατφόρμα ανοιχτού κώδικα που επιτρέπει στους χρήστες να δημιουργούν ροές εργασίας για την εξόρυξη, ενοποίηση και ανάλυση δεδομένων. Λειτουργεί με την χρήση “building blocks” σε ένα περιβάλλον δεν χρειάζεται εκτενής γνώση κώδικα (low-code). Επιπλέον προσφέρει την δυνατότητα ενσωμάτωσης μοντέλων μηχανικής μάθησης καθώς και δυνατότητες οπτικοποίησης που όμως δεν χρησιμοποιούνται για τις ανάγκες του παραδοτέου.

SharePoint OnLine¹⁶

Το SharePoint είναι μια πλατφόρμα συνεργασίας και ενσωματώνεται στο Microsoft Office. Έχει υψηλές δυνατότητες παραμετροποίησης όμως συνήθως αποτελεί πλατφόρμα για την διαχείριση και αποθήκευση αρχείων σε οργανισμούς και επιχειρήσεις. Μέρος του αποτελεί το SharePoint OnLine που συνήθως είναι άμεσα συνδεδεμένο με το Microsoft 365. Ένα από τα πλεονέκτημα του είναι ότι δεν χρειάζεται εγκατάσταση και συντήρηση σε υποδομές του οργανισμού ή της επιχείρησης που το χρησιμοποιεί καθώς όλη η πλατφόρμα παρέχεται ως “Πλατφόρμα ως υπηρεσία” (Platform as a Service-PaaS).

Python¹⁷

Η Python είναι αντικειμενοστραφής, γενικού σκοπού γλώσσα προγραμματισμού που χαρακτηρίζεται από αναγνωσιμότητα και ευκολία στη χρήση. Υποστηρίζει διάφορα πακέτα που επιτρέπουν την επαναχρησιμοποίηση του κώδικα καθώς επίσης και πολλές βιβλιοθήκες και εργαλεία που διευκολύνουν την ανάλυση

¹⁵ [KNIME](#)

¹⁶ [Microsoft 365 - Συνδρομή για εφαρμογές του Office | Microsoft 365](#)

¹⁷ [Python](#)

των δεδομένων. Στο παρόν έργο χρησιμοποιήθηκε η βιβλιοθήκη requests, για τον έλεγχο της εγκυρότητας των διευθύνσεων URL. Επίσης, με την χρήση κάποιων από τις πλέον βασικές βιβλιοθήκες της Python, όπως pandas και NumPy, εξήχθη η πληροφορία σχετικά με το completion όλων των πεδίων που επιστρέφονται από το API, χωρίς να λαμβάνονται υπόψιν οι ειδικές συνθήκες των εκάστοτε πεδίων.

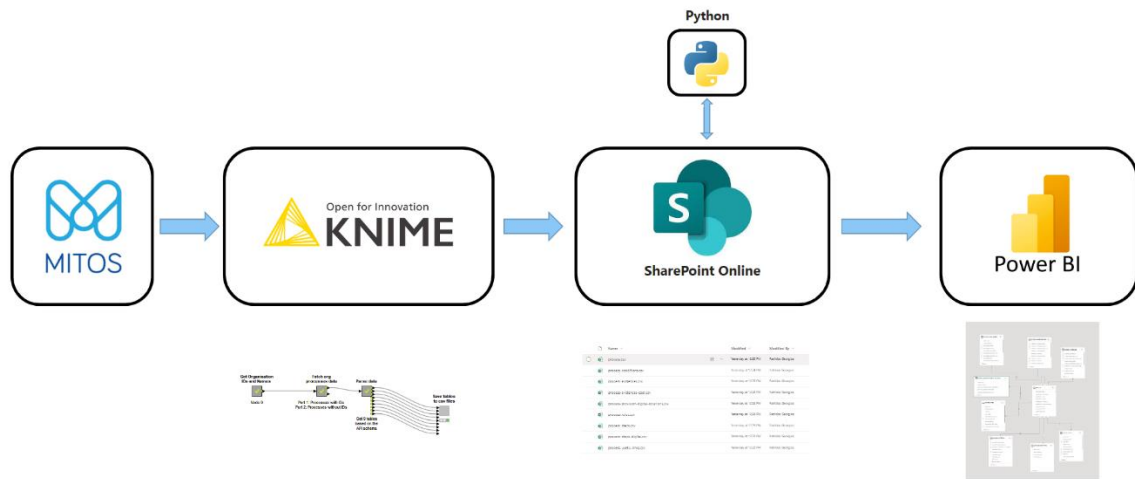
Power BI¹⁸

Το Power BI είναι μια πλατφόρμα επιχειρησιακής ευφυΐας της Microsoft που προσφέρει την δυνατότητα αποθήκευσης, επεξεργασίας και οπτικοποίησης δεδομένων. Επιτρέπει την εισαγωγή δεδομένων από διαφορετικές πηγές παρέχοντας επιπλέον ένα γραφικό περιβάλλον για την επεξεργασία αυτών. Επιπρόσθετα παρέχει την δυνατότητα στους χρήστες να φτιάξουν νέα μοντέλα δεδομένων πάνω στα οποία μπορούν δημιουργήσουν δυναμικές οπτικοποιήσεις και αναφορές.

Η ροή εργασίας υλοποιήθηκε έτσι ώστε να γίνεται η εξόρυξη των δεδομένων από το MitoS API με την χρήση του KNIME, στο οποίο διαμορφώθηκε και η δομή των δεδομένων με τέτοιο τρόπο ώστε να μπορεί να αξιοποιηθεί αργότερα για την οπτικοποίησή τους. Στην συνέχεια, χρησιμοποιήθηκε το SharePoint ως ένα ενδιάμεσο στάδιο για την αποθήκευση των δεδομένων σε μορφή CSV ώστε να μπορούν τα δεδομένα από το MitoS να αξιοποιηθούν και από άλλα εργαλεία. Τέλος, γίνεται χρήση του εργαλείου Power BI προκειμένου να δομηθεί το τελικό μοντέλο των δεδομένων αλλά και οι δυναμικές αναφορές.

Η παρακάτω εικόνα αποτυπώνει τη ροή εργασίας που ακολουθήθηκε για την δημιουργία των δυναμικών αναφορών.

¹⁸ [Data Visualisation | Microsoft Power BI](#)



Εικόνα 3: Ροή εργασίας

Πιο αναλυτικά, ακολουθήθηκαν τα παρακάτω βήματα:

- I. Ανάλυση του Mitos API ώστε να κατανοηθεί η δομή στην οποία επιστρέφει δεδομένα για τις διαδικασίες.
- II. Χρήση του KNIME για την εξόρυξη των διαδικασιών από το Mitos API.
- III. Επεξεργασία και διαμόρφωση των δεδομένων εντός του KNIME ώστε να μπορούν να αποθηκευτούν σε μορφή .CSV αλλά και να αξιοποιηθούν από άλλα εργαλεία.
- IV. Αποθήκευση των .CSV αρχείων στο SharePoint OnLine.
- V. Εξαγωγή επιπλέον πληροφοριών μέσω του Python και αποθήκευση σε .CSV μορφή στο SharePoint OnLine.
- VI. Άντληση των αρχείων με το Power BI.
- VII. Επεξεργασία και τελική διαμόρφωση ενός μοντέλου δεδομένων που επιτρέπει την αποτελεσματική οπτικοποίηση των δεδομένων.
- VIII. Δημιουργία της αναφοράς.

4. Ερευνητική Προσέγγιση

Στην ενότητα αυτή παρουσιάζονται αναλυτικά τα βήματα που ακολουθήθηκαν προκειμένου να αξιολογηθούν τα δεδομένα που συλλέχθηκαν από την πλατφόρμα του ΕΜΔΔ ως προς την ποιότητά τους, λαμβάνοντας υπόψιν τις ανάγκες και τους οργανωτικούς κανόνες του οργανισμού. Οι βασικοί σταθμοί της μεθοδολογίας μας είναι οι ακόλουθοι:

A. Δεδομένου του εύρους και της πολυπλοκότητας του πεδίου που μελετά την ποιότητα των δεδομένων, καθορίστηκε αρχικά ο στόχος της μεθοδολογικής προσέγγισης θέτοντας τα ακόλουθα κριτήρια:

1. Σκοπός (Why): εσωτερικός ποιοτικός έλεγχος (Internal quality monitoring).
Η αξιολόγηση της ποιότητας δεδομένων πραγματοποιείται εξετάζοντας τη διαθέσιμη πληροφορία προς ενημέρωση του data releaser (π.χ. διαχειριστές MITOS) για πιθανές ελλείψεις και αστοχίες.
2. Είδος δεδομένων (What): περιεχόμενο ΕΜΔΔ (περιγραφές διαδικασιών).
Η αξιολόγηση της ποιότητας των δεδομένων θα πραγματοποιηθεί σε επίπεδο περιεχομένου (content).
3. Επίπεδο αξιολόγησης (Data type):
Ο έλεγχος ποιότητας πραγματοποιείται εξετάζοντας:
 - a. πεδία (attributes) που χρησιμοποιούνται στο MITOS για την περιγραφή των διαδικασιών (π.χ. τίτλος, προϋποθέσεις)
 - b. και τιμές των πεδίων αυτών (data values).
4. Τρόπος (How): μέσω ποιοτικών μετρικών (quality KPIS)
5. Μέθοδος (Operation): υπολογιστικά και χειροκίνητα (automated and manual).

Ο έλεγχος της ποιότητας μπορεί να εφαρμοστεί είτε χρησιμοποιώντας μετρικές που μπορούν να υπολογιστούν προγραμματιστικά, είτε χειροκίνητα. Στο πλαίσιο αυτό εφαρμόζονται και οι δύο μέθοδοι.

6. **Stakeholders:** Η προτεινόμενη προσέγγιση απευθύνεται στους διαχειριστές της παρεχόμενης πληροφορίας.

Β. Στη συνέχεια αναπτύχθηκε το πλαίσιο ποιοτικής ανάλυσης και αξιολόγησης που θα ακολουθηθεί καθορίζοντας τις ποιοτικές μετρικές, καθώς και τις αντίστοιχες μονάδες μέτρησης που θα εφαρμοστούν στα δεδομένα. Στην Ενότητα [Μοντέλο ποιοτικής ανάλυσης](#) περιγράφονται αναλυτικά τα επιμέρους δομικά στοιχεία του πλαισίου.

Γ. Τέλος, πραγματοποιείται έλεγχος της ποιότητας των δεδομένων σύμφωνα με τις προκαθορισμένες μετρικές. Για την ανάλυση και την παρουσίαση των αποτελεσμάτων υλοποιήθηκε ένα δυναμικός πίνακας ελέγχου, χρησιμοποιώντας ένα εργαλείο ανάλυσης και οπτικοποίησης δεδομένων.

4.1. Δεδομένα που χρησιμοποιήθηκαν

Τα δεδομένα που εξετάζονται στο παρόν έργο αφορούν τις περιγραφές των δημοσιευμένων διοικητικών διαδικασιών και συγκεκριμένα τα πεδία που τις περιγράφουν και τις πραγματικές τιμές των πεδίων αυτών. Η πληροφορία αυτή συλλέχθηκε από το API του MITOS ακολουθώντας τη διαδικασία που περιγράφηκε στην Ενότητα [Ανάλυση δεδομένων](#).

4.2. Μοντέλο ποιοτικής ανάλυσης

Όπως ήδη έχει προκύψει από τη μελέτη της βιβλιογραφίας, δεν υπάρχει διαθέσιμη κάποια πρότυπη λίστα μετρικών ποιότητας που μπορεί να επαναχρησιμοποιηθεί για την αξιολόγηση της πληροφορίας σε επίπεδο περιεχομένου, καθώς η έννοια της ποιότητας σχετίζεται άμεσα με τα ιδιαίτερα

χαρακτηριστικά του τομέα που εξετάζεται, αλλά και τις ανάγκες και απαιτήσεις του οργανισμού που παρέχει την πληροφορία.

Το συμπέρασμα αυτό σε συνδυασμό με τις καλές πρακτικές που εντοπίστηκαν και τα μεθοδολογικά κριτήρια που ορίστηκαν στο εισαγωγικό τμήμα της ενότητας αυτής μας οδήγησαν στην επιλογή των διαστάσεων που προτείνονται από τις αρχές FAIR (Findability, Accessibility, Interoperability, Reusability) και τις ποιοτικές μετρικές που ορίστηκαν κατά αντιστοιχία από τον ευρωπαϊκό «Οδηγό προς δημοσίευση ποιοτικών δεδομένων»¹⁹. Οι επιλογή των FAIR principles έγινε για τους εξής λόγους:

1. Τα FAIR principles μπορούν να εφαρμοστούν τόσο σε δεδομένα, όσο και σε μεταδεδομένα.
2. Τα FAIR principles υιοθετούνται και προωθούνται από μεγάλους οργανισμούς και την Ευρωπαϊκή Επιτροπή, δίνοντας έμφαση στην εύκολη πρόσβαση και επαναχρησιμοποίηση των δεδομένων.
3. Στον «Οδηγό προς δημοσίευση ποιοτικών δεδομένων» που παρουσίασε η Ευρωπαϊκή Επιτροπή το 2019 , αντιστοιχεί τις αρχές FAIR με τις επικρατέστερες ποιοτικές μετρικές που αναφέρονται στην βιβλιογραφία. Λαμβάνοντας υπόψιν τις παρατηρήσεις αυτές και δεδομένου ότι ο σκοπός του έργου αυτού δεν είναι η αποδελτίωση ή η ομογενοποίηση των υφιστάμενων μετρικών της βιβλιογραφίας, αποφασίσαμε να εξετάσουμε ορισμένες από τις μετρικές αυτές όπως προτείνονται από τον εν λόγω οδηγό.²⁰

Η διαφοροποίηση της δικής μας προσέγγισης αφορά το επίπεδο (granularity level) στο οποίο γίνεται η ανάλυση, εξετάζοντας διαφορετικό

¹⁹ (Data.Europa.Eu Data Quality Guidelines - Publications Office of the EU, n.d.)

²⁰ Ενδέχεται να υπάρχουν διαφοροποιήσεις στους ορισμούς και τις μονάδες μέτρησης ορισμένων μετρικών για την καλύτερη ευθυγράμμιση με τις ανάγκες των δεδομένων μας.

data concept²¹. Πιο συγκεκριμένα, ο ευρωπαϊκός οδηγός παρουσιάζει ένα πλαίσιο ποιότητας δεδομένων/μεταδεδομένων για ένα σύνολο δεδομένων (dataset quality). Σε αντίθεση, το πλαίσιο που προτείνεται στο έργο αυτό επικεντρώνεται αποκλειστικά στην ποιότητα δεδομένων/μεταδεδομένων της πληροφορίας που περιέχεται σε ένα σύνολο δεδομένων (content quality). Για τον ορισμό των μετρικών λάβαμε υπόψιν και τους οδηγούς που παρέχονται από την κοινότητα DAMA-NL²².

4. Τέλος, τα FAIR principles δεν επιβάλλουν κάποιον περιορισμό ως προς την ανοικτή διάθεση των δεδομένων, σε αντίθεση με τα υπόλοιπα πλαίσια αξιολόγησης ανοιχτών κυβερνητικών δεδομένων προϋποθέτουν όλα τα δεδομένα να είναι ανοικτά εξ ορισμού (Higman et al., 2019).²³

Υπενθυμίζεται ότι εστιάζουμε αποκλειστικά σε objective quality metrics, δηλαδή δε λαμβάνουμε υπόψιν το βαθμό ικανοποίησης των χρηστών.

Ο Πίνακας 1 περιγράφει αναλυτικά την κάθε μετρική του πλαισίου (KPI) και παραθέτει την αντίστοιχη μονάδα μέτρησης σε κάθε επίπεδο ανάλυσης (data values, metadata). Για τον καθορισμό των μονάδων μέτρησης λάβαμε υπόψιν το διαστασιολόγιο που προτείνεται από τους (Black, 2020) [Παράρτημα Α: Μετρικές Ανοιχτών Κυβερνητικών Δεδομένων](#)

Πλαίσιο Ανοιχτών Κυβερνητικών Δεδομένων	Μετρικές
--	-----------------

²¹ Το report <https://dama-nl.org/wp-content/uploads/2020/11/How-to-Select-the-Right-Dimensions-of-Data-Quality-v1.1-d.d.-14-Nov-2020.pdf> περιγράφει τα διαφορετικά data concepts (section 2.2, Appendix 5)

²² [Data quality dimension \[Data Management Wiki\]](#)

²³ Δεδομένου ότι το ΜΙΤΟΣ βρίσκεται σε πιλοτικό στάδιο ανάπτυξης και δεν διατίθεται ελεύθερα στο κοινό δεν μπορεί να αξιολογηθεί σύμφωνα με τις διαστάσεις των πλαισίων αυτών.

8 OGD Principles (2007)	completeness, Primary, Timely, Accessible, Machine processable, Non-Discriminatory (available without registration), Non-Proprietary (in terms of format) and License-free),
Sunlight Foundation (2010)	completeness, primacy, timeliness, ease of physical and electronic access, machine readability, nondiscrimination, use of commonly owned standards, licensing, permanence and usage costs.
Open Data Charter (2016)	Open by Default, Timely and Comprehensive, Accessible and usable, Comparable and Interoperable, For Improved Governance and Citizen Engagement, For Inclusive Development and Innovation
Tim Berners-Lee (2006)	5 stars open data

Παράρτημα Β: Μονάδες Μέτρησης Ποιοτικών Μετρικών.

Πίνακας 1: Μοντέλο ποιοτικής αξιολόγησης για τα δεδομένα του ΕΜΔΔ²⁴

Διαστάσεις (Dimensions)	Δείκτης Επίδοσης (KPIs)	Περιγραφή (Description)	Μονάδα Μέτρησης (Measurement)	
			Data (Values)	Metadata
Ευρεσιμότητα (Findability)	Πληρότητα (Completeness)	The data is complete if it includes all values needed to represent an entity (EU Guidelines, 2021)	Number of null values in the mandatory schema ²⁵ fields (%)	N/A

²⁴ Οι περιγραφές των μετρικών παρέχονται στα αγγλικά για λόγους ταύτισης με τη βιβλιογραφία

²⁵ Εξετάζονται τα υποχρεωτικά πεδία τόσο του MITOS schema όσο και του CPSV-AP schema.

	Ανιχνευσιμότητα (Discoverability)²⁶	The degree to which data are easy to find for humans or machines. The discoverability of the data depends on the better description e.g. through usage of keywords (EU Guidelines, 2021).	N/A	1. Missing meta description (boolean) 2. Missing keywords (boolean)
Προσβασιμότητα (Accessibility)	Διαθεσιμότητα (Availability)	Content is available by a human or machine without errors or access restrictions (EU Guidelines, 2021)	1. Check the functionality of URLs ²⁷ (%) 2. Access authorization information is given (Boolean)	N/A
Επανα-χρησιμοποίηση (Reusability)	Ακρίβεια (Accuracy)	The degree of closeness of data values to real values (Black, 2020).	Percentage of accurate allowable values (%).	Percentage of accurate fields (Semantic Distance between metadata fields that support CPSV and fields that are actually included in the dataset.)
	Σχετικότητα (Relevance)	The extent to which data are applicable and helpful for the task at hand (Black, 2020). Thus, the data set should only contain necessary information. (EU Guidelines, 2021).	Number of null values of non-mandatory fields (%). Further elaboration on the usefulness of non-mandatory fields (story). This part is highly dependent on domain experts' perception.	

²⁶ Σύμφωνα με τα EU Guidelines η μετρική αυτή ονομάζεται «Findability», με αποτέλεσμα να έχουμε ίδια ονομασία σε έννοιες που αφορούν διαφορετικό επίπεδο ανάλυσης. Προς αποφυγή συγχύσεων επιλέξαμε να μετονομάσουμε το KPI σε «Discoverability».

²⁷ Γίνεται έλεγχος της λειτουργικότητας των URL που χρησιμοποιούνται σε ορισμένα πεδία ως πληροφορία

	Συνέπεια (Consistency)	The degree to which data values comply with a rule (Black, 2020).	Compliance with business rules ²⁸ (%).	Structure inconsistencies (e.g. missing style guide tags (<h2>))
	Εγκαιρότητα (Timeliness)	Data are timely if they are up to date and represent the actual and current situation (EU Guidelines, 2021)	Last Updated Categorization (dynamically)	N/A
	Αξιοπιστία (Credibility)	Data is considered credible if it is based on trustworthy sources(EU Guidelines, 2021). Credibility includes the concept of authenticity (the truthfulness of origins, attributions, commitments) (Black, 2020)	1. Percentage of processes where source is a credible source (%) e.g. Rule_url "ΔΙΑΥΓΕΙΑ", "ΕΤ", "EURO_LEX" 2. Check the validity of URLs ²⁹	N/A
Διαλειτουργικότητα (Interoperability)	Συμμόρφωση (Conformance)	The degree to which data and metadata is in accordance with accepted standards and regulations (Black, 2020; EU Guidelines, 2021)	Compliance with standards e.g. date format, character encoding (Boolean/story)	Assessment based on CPSV documentation guidelines for conformance statement
	Μηχανική Αναγνωσιμότητα (Machine Readability)	This indicator assesses the extent to which the data and metadata are machine interpretable, i.e. the extent to which they can be understood and handled by automated processes	Evaluation based on based on the 5 star-schema model of Berners-Lee (2006)	N/A

²⁸ A business rule is a criterion used to guide the way operations proceed (e.g. guide day-to-day business activity or make operational business decisions). Business rules can be derived from: data models, business documentation, master/reference data sources (e.g. currency codes).

²⁹ Δειγματοληπτική ανάλυση

		(EU Guidelines, 2021).		
--	--	------------------------	--	--

4.3. Τεχνική Υλοποίηση και Αποτελέσματα

Η ενότητα αυτή αφορά το τρίτο και τελευταίο στάδιο της μεθοδολογίας μας, στο οποίο πραγματοποιείται ο έλεγχος της ποιότητας της παρεχόμενης πληροφορίας του ΕΜΔΔ, σύμφωνα με τις μετρικές που ορίστηκαν και παρουσιάστηκαν στον Πίνακα 1. Η τεχνική υλοποίηση του σταδίου αυτού περιλαμβάνει δύο βήματα:

4.3.1. Επιλογή των πεδίων στα οποία θα εφαρμοστούν οι προκαθορισμένες μετρικές.

Επισημαίνεται ότι στο συγκεκριμένο παραδοτέο δίνεται προτεραιότητα³⁰ στην αξιολόγηση των μετρικών που επιτρέπουν την ανάλυσή τους προγραμματιστικά (Πίνακας 1 / είδος μονάδας μέτρησης: %,) και σε επίπεδο πραγματικών τιμών “data(values)”.

Στον Πίνακα 2 περιγράφεται το είδος των πεδίων του API που αξιολογήθηκαν για κάθε μετρική. Τα πεδία που ανήκουν στις μετρικές completeness, availability και consistency παρουσιάζονται αναλυτικά στο [Παράρτημα Γ: Πεδία ανά μετρική](#).

Πίνακας 2: Πεδία του ΕΜΔΔ που αξιολογήθηκαν ανά μετρική

Μετρικές	Πεδία που αξιολογήθηκαν	Τύπος Υπολογισμού (Formula)
Πληρότητα	1. Υποχρεωτικά πεδία στο schema του	[Αριθμός μη-συμπληρωμένων

³⁰ Σύμφωνα με τις άμεσες ανάγκες του ΕΜΔΔ, όπως αυτές διαμορφώθηκαν από τους ειδικούς πεδίου του οργανισμού

(Completeness)	ΜΙΤΟΣ (βάσει εγχειριδίου συντακτών) 2.Υποχρεωτικά πεδία του CPSV-AP	πεδίων / Αριθμός συνολικών υποχρεωτικών πεδίων] Κλίμακα: [0%-100%]
Διαθεσιμότητα (Availability)	Πεδία που η τιμή τους αναμένεται να είναι URL π.χ. url αίτησης (process_related_url)	[Αριθμός λειτουργικών (valid) url ³¹ / Αριθμός συνολικών url] Κλίμακα: [0%- 100%]
Ακρίβεια (Accuracy)	Όλα τα πεδία που παρέχονται από το API ως προς την αναμενόμενη μορφή σύνταξής τους (format).	Υπολογίζεται αυτόματα με στατιστικά κριτήρια από το εργαλείο ανάλυσης δεδομένων. Κλίμακα: [0%- 100%]
Συνέπεια (Consistency)	Πεδία τα οποία καθορίζονται από λογικούς και θεσμικούς κανόνες όπως αυτοί περιγράφονται στο διαθέσιμο εγχειρίδιο των συντακτών.	Για τον τρόπο υπολογισμού της μετρικής αυτής εφαρμόζονται λογικοί κανόνες οι οποίοι εξαρτώνται κάθε φορά από τον κανόνα που έχει οριστεί για ένα πεδίο (βλ. Παράρτημα Γ: Πεδία ανά μετρική).
Εγκαιρότητα (Timeliness)	Πεδίο «last updated». ³²	
Αξιοπιστία (Credibility)	Πεδίο «rule_url» ως προς τον κανόνα: «ΔΕΝ πρέπει να εισάγονται σύνδεσμοι που προέρχονται από άλλες νομικού περιεχομένου ιστοσελίδες εκτός των: Εθνικό Τυπογραφείο ή στη Διαύγεια ή στον ιστότοπο EUR-Lex»	[Αριθμός των rule urls με αποδεκτή προέλευση/ Συνολικός Αριθμός συμπληρωμένων πεδίων «rule_url»] Κλίμακα: [0%- 100%]

³¹ Για τον έλεγχο λειτουργικότητας των url χρησιμοποιήθηκε η βιβλιοθήκη *requests* της Python. Ως valid urls θεωρούνται όσα έχει ανατεθεί τύπος Μηνύματος: [website exists]

³² Το πεδίο "last_updated" παρέχει την ημερομηνία τελευταίας ενημέρωσης μιας διαδικασίας και σχετίζεται με οποιαδήποτε αλλαγή που μπορεί να έγινε στην περιγραφή μιας διαδικασίας. Παρά το ότι δεν παρέχει ουσιαστική πληροφορία, στην παρούσα φάση είναι το μοναδικό πεδίο από αυτά που παρέχει το API και μπορεί να συσχετιστεί με την μετρική timeliness.

4.3.2. Δημιουργία δυναμικού dashboard για την ανάλυση και την οπτικοποίηση των αποτελεσμάτων μας

Η ανάλυση των δεδομένων μπορεί να επιβεβαιώσει ή να απορρίψει υποθέσεις καθώς και να δώσει κατεύθυνση για νέες υποθέσεις ως προς την ποιότητα των δεδομένων που βρίσκονται στο ΕΜΔΔ. Επιπρόσθετα η οπτικοποίηση των αποτελεσμάτων βοηθάει στην καλύτερη κατανόηση τους.

Στο πλαίσιο της ανάλυσης και αξιολόγησης των δεδομένων του ΕΜΔΔ συγκεντρώθηκαν συνολικά πληροφορίες για 3033 διαδικασίες οι οποίες αναλύονται στο εργαλείο PowerBI με βάση τις μετρικές και τις μονάδες μέτρησης που παρουσιάζονται στον Πίνακα 2. Το αποτέλεσμα είναι η δημιουργία μίας δυναμικής αναφοράς, η οποία περιλαμβάνει ξεχωριστές σελίδες ανά μετρική. Σε κάθε σελίδα παρουσιάζεται ο βαθμός απόδοσης των πεδίων που εξετάζονται και ανήκουν στην εκάστοτε μετρική. Στην Εικόνα 4 απεικονίζεται η αρχική σελίδα της δυναμικής αναφοράς η οποία περιλαμβάνει μία λίστα πλοήγησης των περιεχομένων της καθώς και ένα συνοπτικό εγχειρίδιο με τα πεδία που εξετάζονται σε κάθε μετρική.

Σημειώνεται ότι τα δεδομένα που αναλύονται έχουν συλλεχθεί από το Mitos API με ημερομηνία τελευταίας ανανέωσης την 26η Ιουλίου 2023.

The screenshot shows the MITOS system interface. At the top left, there are logos for MITOS and the International Hellenic University, along with the date 26/7/2023 and the number 3033, which represents the 'Number of Processes'. A navigation pane on the left lists various dashboard options, with 'Navigation Pane' selected. The main content area on the right is titled 'Pages info' and contains detailed information about the 'Workflow information' and 'Accuracy Dashboard'. It lists various metrics such as 'Accuracy-1', 'Accuracy-2', 'Accuracy-3', 'All Fields Completeness', 'Mandatory Fields Completeness Dashboard', 'Completeness-1' through 'Completeness-7', 'Relevance', 'Timeliness', 'Credibility', 'Availability-1', 'Availability-2', and 'Consistency-1' through 'Consistency-5'. Each metric is accompanied by a list of associated data fields in brackets. For example, 'Accuracy-1' is associated with [process, conditions, evidences].

Εικόνα 4: Σελίδα πλοήγησης της δυναμικής αναφοράς

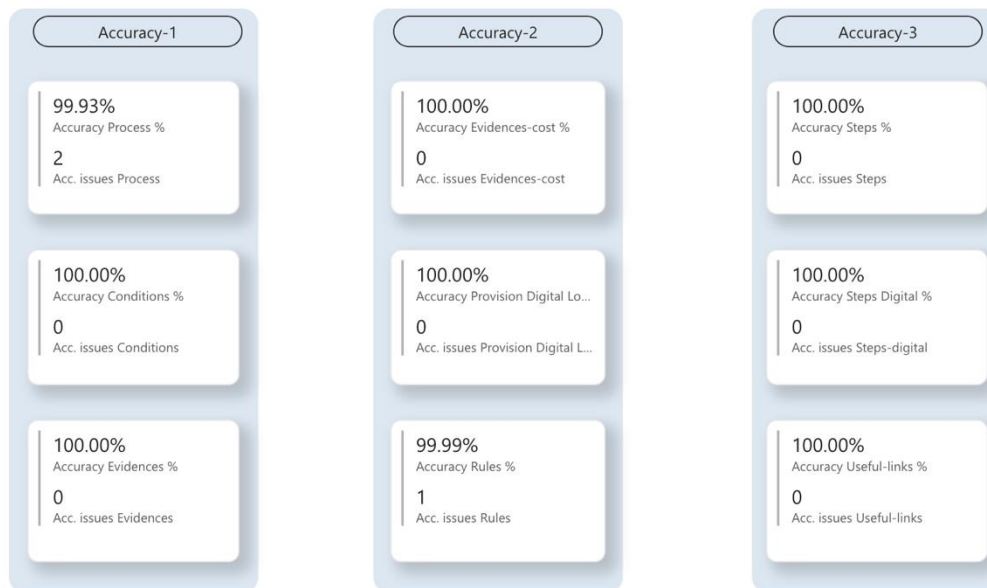
Στη συνέχεια, παρουσιάζονται αντιπροσωπευτικά ευρήματα που έχουν επιλεγεί από την αναφορά για κάθε μετρική μαζί με συμπεράσματα όπως αυτά συνάγονται από την ανάλυση των δεδομένων. Η πλήρης αναφορά είναι άμεσα διαθέσιμη μέσω συνδέσμου³³.

➤ Ακρίβεια (Accuracy)

Η πρώτη σελίδα ονομάζεται “Accuracy Dashboard” και αποτελεί μία συνοπτική παρουσίαση όλων των αποτελεσμάτων που αφορούν την μετρική *Accuracy*, ακολουθώντας τις κατηγορίες του σχήματος (schema) του MITOS API, όπως περιγράφεται στο Κεφάλαιο 3. Πιο συγκεκριμένα με βάση την ανάλυση που παρουσιάζεται στην Εικόνα 5, στην κατηγορία Process του σχήματος

³³ [Δυναμική Αναφορά \(Report\)](#)

εμφανίζονται 2 issues και ποσοστό ακρίβειας 99.93%, ενώ στην κατηγορία Rules 1 issue με ποσοστό ακρίβειας 99.99%. Σε όλες υπόλοιπες κατηγορίες δεν εμφανίζεται κάποιο issue, κατ' επέκταση η πληρότητα διαμορφώνεται στο 100%.



Εικόνα 5: Έλεγχος Ακρίβειας (Accuracy)

Στη συνέχεια, τα συγκεντρωτικά αποτελέσματα αναλύονται περαιτέρω σε υποσελίδες, όπου παρουσιάζονται αναλυτικά τα Errors που εξήχθησαν για κάθε περίπτωση. Αναλυτικότερα, κατά τη φόρτωση των δεδομένων γίνεται για κάθε πεδίο μια αυτόματη αναγνώριση του τύπου του με βάση την κατανομή των τιμών³⁴. Για παράδειγμα, αν ένα πεδίο περιλαμβάνει κυρίως αριθμούς τότε αναγνωρίζεται ως τύπος "Number". Σε περίπτωση που υπάρξει κείμενο σε αυτό το πεδίο τότε δημιουργείται ένα Error με αιτιολογία "DataFormat.Error".

³⁴ <https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-data-types>

Αυτό γίνεται καλύτερα αντιληπτό στην Εικόνα 6, όπου αναλύονται τα 2 Errors της κατηγορίας Process. Παρατηρώντας το πρώτο Error, το πεδίο "org_owner" έχει αναγνωριστεί από το σύστημα ως "Number", όμως στην προκειμένη περίπτωση έχει συμπληρωθεί λανθασμένα, όπως φαίνεται στην στήλη "Detail". Επίσης, καταγράφονται τα πεδία "process_id" και "title_el", δίνοντας την δυνατότητα στον χρήστη της επαλήθευσης και διόρθωσης του Error που αναγνωρίστηκε. Τέλος, να σημειωθεί ότι τα Errors προσμετρούνται ως Null/Blank κατά την ανάλυση, για να μην αφαιρεθεί πολύτιμη πληροφορία στην συνέχεια.

process_id	title_el	Column name	Reason	Message	Detail
449862	Μεταβίβαση σύνταξης στον επώνυμο σύζυγο λογοτέχνη - καλλιτέχνη	org_owner	DataFormat.Error	We couldn't convert to Number.	"83842, 83842"
879661	Πρόσληψη κεραισίων εκπαιδευτικών με σχέση εργασίας Ι.Δ.Ο.Χ. στη Δημόσια Γενική Πρωτοβάθμια και Δευτεροβάθμια Εκπαίδευση	estimated_implementation_time	DataFormat.Error	We couldn't convert to Number.	Η διαδικασία υλοποιείται εντός του εκάστοτε διδακτικού έτους.

Εικόνα 6: Λάθη Ακρίβειας που αφορούν διαδικασίες

➤ Πληρότητα (Completeness)

Κατά την ανάλυση την μετρικής Completeness εξετάζεται η πληρότητα μόνο των υποχρεωτικών (mandatory) πεδίων, όπως αυτά ορίζονται από το CPSV schema και το εγχειρίδιο των συντακτών. Αρχικά τα αποτελέσματα παρουσιάζονται συγκεντρωτικά χωρισμένα ανά κατηγορία του σχήματος (Εικόνα 7). Ειδικότερα, για κάθε πεδίο υπάρχουν 3 βασικές μετρικές. Η κύρια μετρική που παρουσιάζεται με διάγραμμα "gauge" εξετάζει το ποσοστό πληρότητας του πεδίου. Η δεύτερη μετρική παραθέτει τον απόλυτο αριθμό των τιμών που δεν έχουν συμπληρωθεί και η τελευταία το ποσοστό αυτών με βάση το σύνολο.

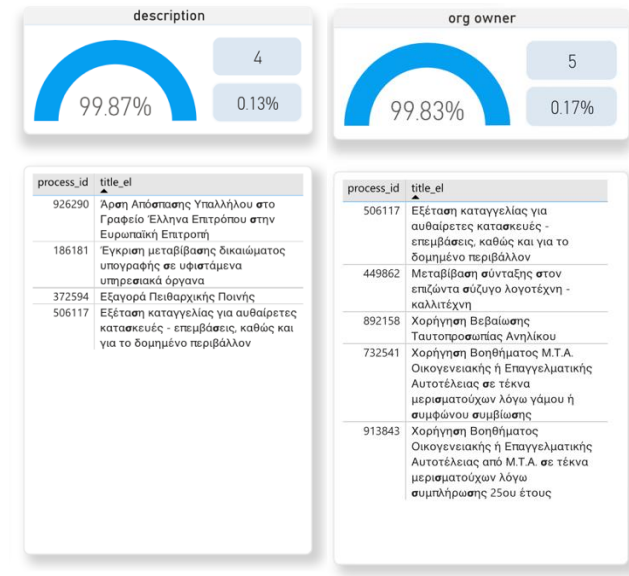
Για την εξαγωγή των ποσοστών λαμβάνονται υπόψιν οι εκάστοτε ειδικές συνθήκες. Για παράδειγμα στην περίπτωση του "org_owner" ο αριθμητής και ο

παρανομαστής προσαρμόζονται αφαιρώντας από το σύνολο των διαδικασιών (3033) τις διαδικασίες που έχουν τιμή "1" για το πεδίο "org_owner_is_private".



Εικόνα 7: Έλεγχος Πληρότητας (Completeness) των υποχρεωτικών πεδίων

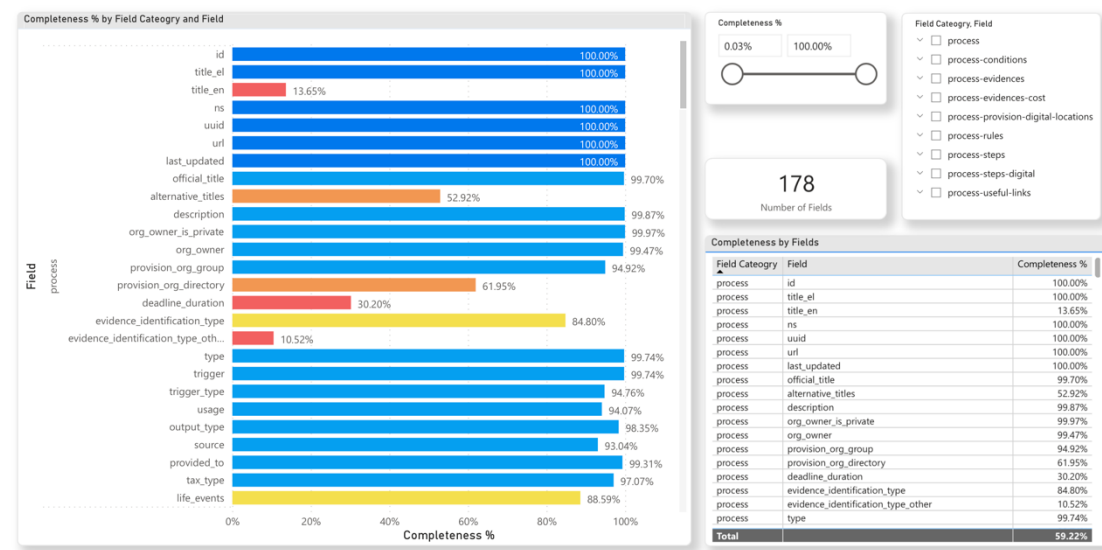
Όπως και στην περίπτωση του Accuracy, τα αποτελέσματα του συγκεντρωτικού Dashboard αναλύονται περαιτέρω σε 7 υπό-σελίδες, όπου πέρα από τις μετρικές που αναφέρθηκαν προηγουμένως παρουσιάζονται αναγνωριστικά στοιχεία της διαδικασίας, ώστε να μπορεί να διενεργηθεί πρόσθετος έλεγχος. Παραδείγματα της αναλυτικής καταγραφής του Completeness αποτυπώνονται στην Εικόνα 8. Ενδεικτικά, το πεδίο "description" δεν έχει συμπληρωθεί 4 φορές, επομένως το Completeness υπολογίζεται έχοντας ως αριθμητή το σύνολο των διαδικασιών μείον 4 που είναι κενές και σαν παρανομαστή το σύνολο των διαδικασιών.



Εικόνα 8: Completeness των υποχρεωτικών πεδίων "description" & "org_owner"

Για την καλύτερη εποπτεία της πληροφορίας του MITOS API η μετρική Completeness εφαρμόστηκε και ως προς σύνολο των πεδίων που επιστρέφονται από το API (στην προκειμένη περίπτωση 178 πεδία) χωρίς να λαμβάνονται υπόψη ειδικές συνθήκες που μπορεί να εφαρμόζονται στα εκάστοτε πεδία (π.χ. το πεδίο "org_owner" δεν μπορεί να συμπληρωθεί όταν είναι το πεδίο "org_owner_private" έχει τιμή "1") (Εικόνα 9).

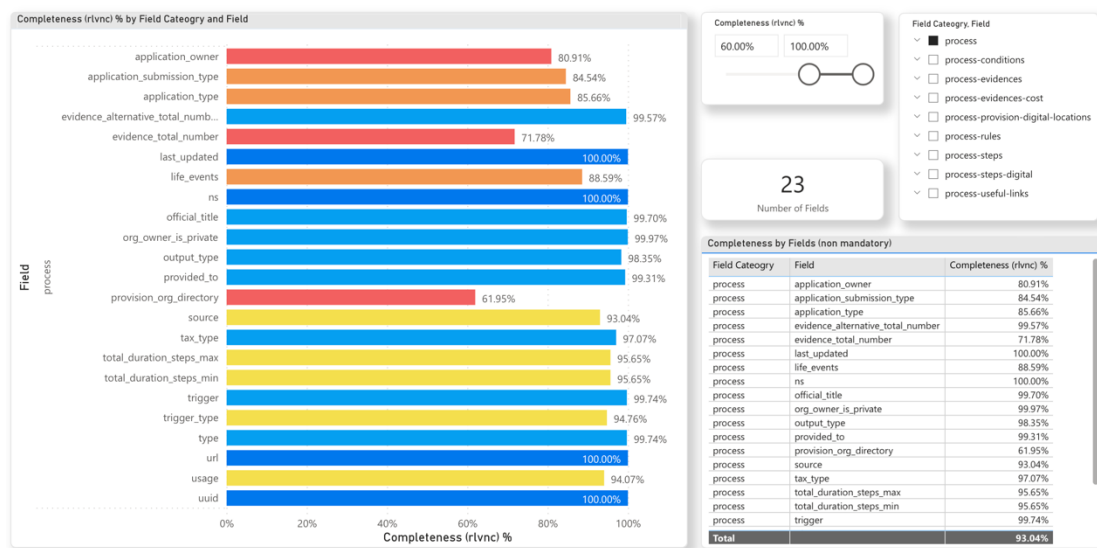
Κάποια από τα πλεονεκτήματα της συγκεκριμένης σελίδας, είναι η δυνατότητα του χρήστη να εξετάσει το Completeness ανά κατηγορία του σχήματος, να περιορίσει το πεδία που εξετάζονται με βάση το ποσοστό πληρότητας τους καθώς και να δει το συνολικό ποσοστό πληρότητας είτε όλων των πεδίων (59,22% στο παράδειγμα μας), είτε μίας κατηγορίας του σχήματος που θα επιλέξει.



Εικόνα 9: Πληρότητα (Completeness) όλων των πεδίων χωρίς ειδικές συνθήκες

➤ Σχετικότητα (Relevance)

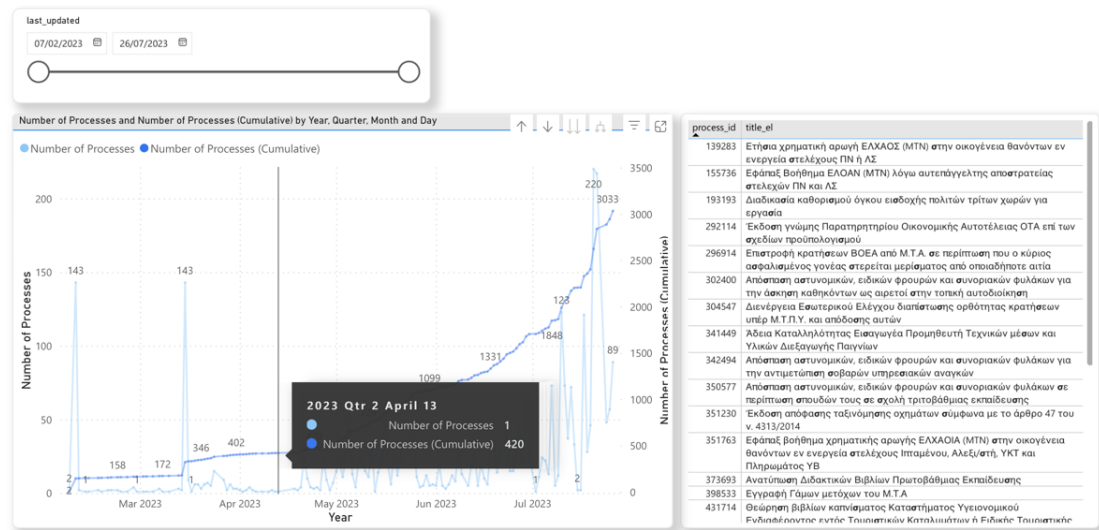
Η ανάλυση της μετρικής Relevance, ακολουθεί παρόμοια λογική με το Completeness όλων των πεδίων που παρουσιάστηκε παραπάνω, εξετάζοντας μόνο τα non-Mandatory πεδία χωρίς να εμπεριέχονται οι ειδικές συνθήκες τους. Αναλυτικότερα στο παράδειγμα που παρουσιάζεται στην Εικόνα 10, παρουσιάζονται τα στατιστικά των 23 πεδίων από την κατηγορία process του σχήματος του MITOS API, τα οποία έχουν πληρότητα μεγαλύτερη του 60%. Με βάση αυτά τα πεδία η πληρότητα των process είναι 93.04%, τονίζοντας πως αυτά τα πεδία παρέχουν αρκετά σχετική και χρήσιμη πληροφορία προς τους χρήστες.



Εικόνα 10: Έλεγχος πληρότητας των μη υποχρεωτικών πεδίων των διαδικασιών (αφορά τη μετρική Relevance)

➤ Εγκαιρότητα (Timeliness)

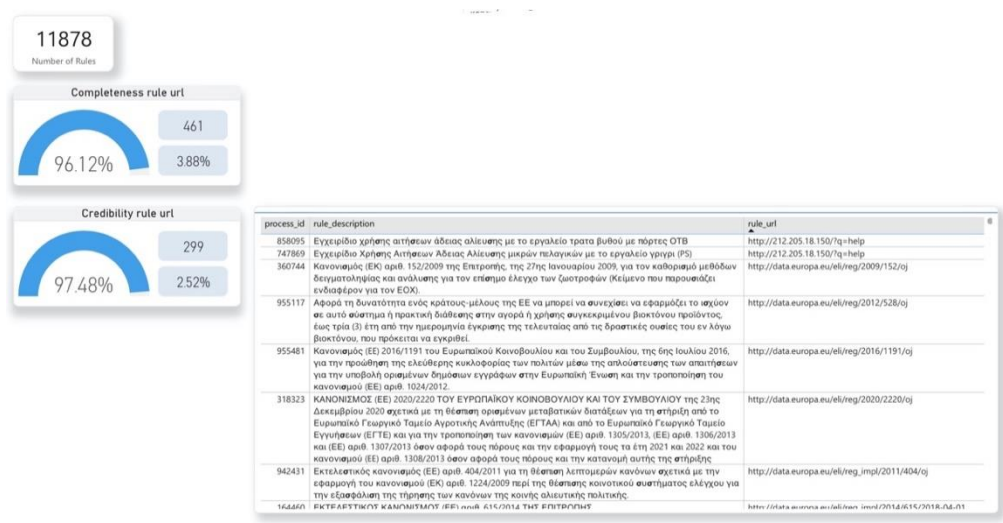
Η ανάλυση της μετρικής Timeliness, πραγματοποιείται σύμφωνα με το πεδίο "last_updated". Μέσω ενός γραφήματος τύπου "line chart" παρουσιάζεται για κάθε μέρα ο αριθμός των διαδικασιών που έχουν ενημερωθεί τελευταία φορά στο ΜΙΤΟΣ (Number of processes) ενώ απεικονίζεται και η προοδευτική αύξηση αυτού του αριθμού των διαδικασιών στην πορεία του χρόνου (Number of Processes cumulative). Συγκεκριμένα, στο παράδειγμα της Εικόνας 11, υπάρχει 1 διαδικασία που ενημερώθηκε τελευταία φορά στις 13 Απριλίου και συνολικά 420 που έχουν ενημερωθεί τελευταία φορά πριν από αυτή την ημερομηνία. Επιπροσθέτως, υπάρχει το φίλτρο "last_updated" που επιτρέπει τον χρήστη να εστιάσει σε μία συγκεκριμένη χρονική περίοδο καθώς και ένας πίνακας που παραθέτει τις διαδικασίες για την επιλεγμένη περίοδο.



Εικόνα 11: Έλεγχος εγκαιρότητας (Timeliness) βάσει του πεδίου "last_updated"

➤ Αξιοπιστία (Credibility)

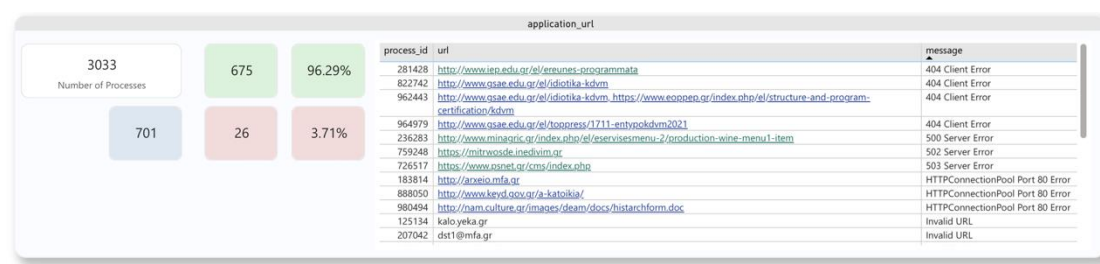
Για την ανάλυση της μετρικής του Credibility εξετάζεται το πεδίο "rule_url", όπου με βάση το manual των συντακτών οι αποδεκτές τιμές συμπλήρωσης του πεδίου είναι οι εξής: ".et.gr/", "diavgeia.gov.gr/", "eur-lex.europa.eu/". Συγκεκριμένα στο παράδειγμα μας (Εικόνας 12), 299 URLs οι τιμές δεν βρίσκονται μέσα σε αυτό το εύρος επιτρεπτών τιμών. Σε αυτό το σημείο πρέπει να τονιστεί πως για την εξαγωγή των ποσοστών στον πίνακα "Credibility rule url" αφαιρούνται οι κενές τιμές του πεδίου δηλαδή τα 461 κενά που έχουν εντοπιστεί.



Εικόνα 12: Έλεγχος αξιοπιστίας (Credibility) βάσει του πεδίου «rule_url»

➤ Εγκυρότητα (Validity)

Στην παρακάτω Εικόνα 12, αναλύεται το Validity του πεδίου “application_related_url” δηλαδή της εγκυρότητας του πεδίου. Ο αριθμός στο μπλε πλαίσιο παραθέτει ότι υπάρχουν 701 URLs στο σύνολο, ενώ στο πράσινο και στο κόκκινο παρουσιάζονται τα URLs που είναι έγκυρα και μη αντίστοιχα. Τα εκάστοτε ποσοστά υπολογίζονται με βάση τον αριθμό των 701 URLs που υπάρχουν στο σύνολο. Στον πίνακα που συνοδεύει τα αποτελέσματα παρουσιάζονται τα URLs που δεν είναι έγκυρα καθώς και το μήνυμα σφάλματος που λαμβάνει το σύστημα. Ίδια ανάλυση πραγματοποιείται σε όσα πεδία έχουν URLs στις τιμές τους.



Εικόνα 13: Έλεγχος εγκυρότητας (Validity) των διευθύνσεων URL των διαδικασιών

➤ **Συνέπεια (Consistency)**

Τέλος, η μετρική που αφορά το Consistency εξετάζεται μέσα από τον έλεγχο κανόνων (Business Rules), όπως αυτοί παρουσιάζονται στο εγχειρίδιο των συντακτών. Κάποιοι από αυτούς τους κανόνες εξετάζουν τις τιμές που έχουν συμπληρωθεί μόνο σε ένα πεδίο. Αντιπροσωπευτικό είναι το πρώτο παράδειγμα της Εικόνας 14, όπου στο πεδίο "provided_languages" η γλώσσα "Ελληνικά" είναι συμπληρωμένη 708 φορές, γεγονός που δεν επιτρέπεται με βάση το manual. Υπάρχουν και κάποιοι κανόνες όπου απαιτείται ο συνδυαστικός έλεγχος πεδίων, όπως στο δεύτερο παράδειγμα που παρατίθεται παρακάτω. Στο συγκεκριμένο εξετάζονται οι περιπτώσεις που το πεδίο "evidence_cost_type" είναι συμπληρωμένο με την τιμή "e-Παράβολο" και το πεδίο "evidence_cost_code", δηλαδή ο κωδικός παράβολου, είναι λανθασμένα κενό.

process_id	title_el	provided_language
103345	Προκαταβολή έναντι εφάπαξ βοηθήματος ΕΛΟΑΝ (MTN) εν ενεργεία στελεχών ΠΝ και ΛΣ	Ελληνικά
103609	Επιλογή Διευθυντών και Προισταμένων Προτύπων και Πειραματικών Σχολείων	Ελληνικά
105058	Προσωρινή Διασυνοριακή Παροχή Υπηρεσιών Επόπτη Δημόσιας Υγείνης	Αγγλικά,Ελληνικά
107962	Έκδοση Άδειας Κυκλοφορίας Φορτηγού Ιδιωτικής Χρήσης (ΦΙΧ) σε Κατηγορίες Προσώπων Μη	Ελληνικά

process_id	evidence_cost_description	evidence_cost_type	evidence_cost_code
105154	1η κατηγορία Σταθμών Υπεραστικών Λεωφορείων: 100€	e-Παράβολο	
105154	2η κατηγορία Σταθμών Υπεραστικών Λεωφορείων: 150€	e-Παράβολο	
105154	3η κατηγορία Σταθμών Υπεραστικών Λεωφορείων: 200€	e-Παράβολο	
105154	4η κατηγορία Σταθμών Υπεραστικών Λεωφορείων: 300€	e-Παράβολο	

Εικόνα 14: Έλεγχος Consistency μέσω Business Rules

5. Συμπεράσματα

5.1. Σύντομη Επισκόπηση

Το συγκεκριμένο παραδοτέο αποτελεί το πρώτο και σημαντικότερο βήμα για τον τελικό στόχο του έργου που αφορά τη βελτίωση της ποιότητας των περιγραφών των διαδικασιών του Εθνικού Μητρώου Διοικητικών Διαδικασιών (ΕΜΔΔ). Σε αυτό το πλαίσιο, έχοντας ως κύριο άξονα την εξέταση της ποιότητας των δεδομένων που παρέχονται από το ΕΜΔΔ 1) πραγματοποιήθηκε μελέτη των δεδομένων που παρέχονται από το API του ΕΜΔΔ ως προς τη δομή και το περιεχόμενό τους. Παράλληλα, 2) μελετήθηκαν καλές πρακτικές για την αξιολόγηση της ποιότητας των κυβερνητικών δεδομένων που προτείνονται από την διεθνή βιβλιογραφία, προκειμένου να γίνει πλήρως κατανοητό το πρόβλημα που καλούμαστε να αντιμετωπίσουμε. Τα αποτελέσματα των δύο μελετών μας οδήγησαν στα εξής συμπεράσματα και παρατηρήσεις:

Μελέτη δεδομένων ΕΜΔΔ:

- Κατά την ανάλυση και διαχείριση των δεδομένων:
 - εντοπίστηκαν ελλείψεις μοναδικών αναγνωριστικών id για βήματα/δικαιολογητικά/κανόνες
 - αντιμετωπίστηκε δυσκολία στην κατανόηση ορισμένων πεδίων και παρατηρήθηκε έλλειψη σχετικών περιγραφών στο εγχειρίδιο των συντακτών
- Κατά την αξιολόγηση των δεδομένων:
 - εντοπίστηκαν συντακτικές ασυνέπειες (formatting issues) σε διαδικασίες
 - ασυνέπειες με τα business rules του φορέα
 - ελλείψεις ως προς την πληρότητα των πεδίων (τόσο στα υποχρεωτικά όσο και στα προαιρετικά πεδία)
 - Ελλείψεις χρήσιμων πεδίων για την αξιολόγηση μετρικών (π.χ. timeliness)

Μελέτη καλών πρακτικών:

- ο Μεγάλο εύρος διαθέσιμης πληροφορίας και αλληλοεπικαλύψεις των εννοιών
- ο Περιορισμένη πληροφορία ως προς την αξιολόγηση κυβερνητικών δεδομένων σε επίπεδο περιεχομένου

Τέλος, 3) αναπτύξαμε τη μεθοδολογική μας προσέγγιση σύμφωνα πάντα με τις άμεσες ανάγκες του ΕΜΔΔ.

Για την ανάλυση και αξιολόγηση της ποιότητας των δεδομένων υιοθετήθηκαν ως βασικές διαστάσεις οι αρχές FAIR (Findability, Accessibility, Interoperability, Reusability) και εφαρμόστηκαν οι αντίστοιχες ποιοτικές μετρικές (π.χ. πληρότητα, συνέπεια και αξιοπιστία) όπως παρουσιάζονται στον ευρωπαϊκό «Οδηγό προς δημοσίευση ποιοτικών δεδομένων». Οι μετρικές αυτές ορίζονται μοναδικά και προτείνονται συγκεκριμένες μονάδες μέτρησης για την αξιολόγησή τους.

Στο συγκεκριμένο παραδοτέο παρουσιάζεται η απόδοση των προγραμματιστικά μετρήσιμων μετρικών (accuracy, availability, completeness, consistency, credibility, timeliness) μέσα από την δημιουργία μίας δυναμικής αναφοράς³⁵ χρησιμοποιώντας εργαλεία ανάλυσης εξόρυξης και ανάλυσης και δεδομένων.

5.2. Κατεύθυνση επόμενων βημάτων

Για το επόμενο στάδιο του έργου θέτονται οι εξής στόχοι:

1. Αξιολόγηση των μετρικών που δεν μετριοούνται προγραμματιστικά
2. Αξιολόγηση της ποιότητας των μετρικών σε επίπεδο μεταδεδομένων π.χ. ως προς τη μετρική accuracy θα εξεταστεί το ποσοστό των πεδίων του MITOS API που συμφωνούν σημασιολογικά με πεδία του CPSV
3. Εκτίμηση της συνολικής εικόνας του ΕΜΔΔ ως προς την ποιότητα της παρεχόμενης πληροφορίας του

³⁵ [Microsoft Power BI](#)

4. Ανάπτυξη προτάσεων σχετικά με ενέργειες και πρακτικές που μπορούν να εφαρμοστούν για την περαιτέρω βελτίωση της ποιότητας των δεδομένων του ΕΜΔΔ βάσει των συμπερασμάτων και των παρατηρήσεων που προκύπτουν από τη συλλογή, τη διαχείριση και την αξιολόγησή τους.

6. Βιβλιογραφία

- Alexopoulos, C., Saxena, S., Rizun, N., & Shao, D. (2023). A framework of open government data (OGD) e-service quality dimensions with future research agenda. *Records Management Journal*, 33(1), 20–32. <https://doi.org/10.1108/RMJ-06-2022-0017>
- Andrade, M. C., Cunha, R. O. da, Figueiredo, J., & Baptista, A. A. (2021). Do the European Data Portal Datasets in the Categories Government and Public Sector, Transport and Education, Culture and Sport Meet the Data on the Web Best Practices? *Data*, 6(8), 94. <https://doi.org/10.3390/data6080094>
- Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, 41(3). <https://doi.org/10.1145/1541880.1541883>
- Black, A. , N. P. van. (2020). Black, A., Nederpelt, P. van. (2020). *Dimensions of Data Quality Dimensions. Research paper. DAMA-NL. - Presearch.* Dimensions of Data Quality Dimensions. Research Paper. DAMA-NL. [https://presearch.com/search?q=Black%2C%20A.%2C%20Nederpelt%2C%20P.%20van.%20\(2020\).%20Dimensions%20of%20Data%20Quality%20Dimensions.%20Research%20paper.%20DAMA-NL.](https://presearch.com/search?q=Black%2C%20A.%2C%20Nederpelt%2C%20P.%20van.%20(2020).%20Dimensions%20of%20Data%20Quality%20Dimensions.%20Research%20paper.%20DAMA-NL)
- Bouchelouche, K., Ghomari, A. R., & Zemmouchi-Ghomari, L. (2022). Enhanced analysis of Open Government Data: Proposed metrics for improving data quality assessment. *2022 5th International Symposium on Informatics and Its Applications (ISIA)*, 1–6. <https://doi.org/10.1109/ISIA55826.2022.9993482>

- Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. *IEEE Access*, 7, 24634–24648. <https://doi.org/10.1109/ACCESS.2019.2899751>
- Data.europa.eu data quality guidelines* - Publications Office of the EU. (n.d.). Retrieved July 20, 2023, from <https://op.europa.eu/da/publication-detail/-/publication/b601d9cc-b3c0-11ec-9d96-01aa75ed71a1>
- datos.gob.es. (2022). *Technical Standards to achieve Data Quality*. <https://datos.gob.es/en/blog/technical-standards-achieve-data-quality>
- Diamond, I., & Chisholm, A. (2020). *The Government Data Quality Framework*. <https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework>
- EU Guidelines. (2021). *Data.europa.eu Data Quality Guidelines - August 2021*. <https://doi.org/10.2830/879764>
- EU Open Data Goldbook. (2018). *Data providers' guide | data.europa.eu*. <https://data.europa.eu/en/training/data-providers-guide>
- Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: The intersections of research data management, FAIR and Open. *Insights: The UKSG Journal*, 32. <https://doi.org/10.1629/uksg.468>
- Ijab, M. T., Ahmad, A., Kadir, R. A., & Hamid, S. (2017). *Towards Big Data Quality Framework for Malaysia's Public Sector Open Data Initiative* (pp. 79–87). https://doi.org/10.1007/978-3-319-70010-6_8
- Jansen, A., & Ølnes, S. (2016). The nature of public e-services and their quality dimensions. *Government Information Quarterly*, 33(4), 647–657. <https://doi.org/10.1016/j.giq.2016.08.005>
- Kučera, J., Chlapek, D., & Nečaský, M. (2013). *Open Government Data Catalogs: Current Approaches and Quality Perspective* (pp. 152–166). https://doi.org/10.1007/978-3-642-40160-2_13
- Lacagnina, C., David, R., Kuusniemi, M. E., Biehlmaier, O., Wright, L., & Dennis, R. (2023). *TOWARDS A DATA QUALITY FRAMEWORK FOR EOSC*.

- Lämmel, P., Dittwald, B., Bruns, L., Tcholtchev, N., Glikman, Y., Cuno, S., Flügge, M., & Schieferdecker, I. (2020). Metadata Harvesting and Quality Assurance within Open Urban Platforms. *Journal of Data and Information Quality*, 12(4), 1–20. <https://doi.org/10.1145/3409795>
- Moraga, C., Moraga, M. Á., Calero, C., & Caro, A. (2009). SQuaRE-Aligned Data Quality Model for Web Portals. *2009 Ninth International Conference on Quality Software*, 117–122. <https://doi.org/10.1109/QSIC.2009.23>
- Moreno, A., Molano-Pulido, J., Gomez-Morantes, J. E., & Gonzalez, R. A. (2022). ADACOP: A Big Data Platform for Open Government Data. *15th International Conference on Theory and Practice of Electronic Governance*, 369–375. <https://doi.org/10.1145/3560107.3560310>
- Nikiforova, A. (2020). Definition and Evaluation of Data Quality: User-Oriented Data Object-Driven Approach to Data Quality Assessment. *Baltic Journal of Modern Computing*, 8(3). <https://doi.org/10.22364/bjmc.2020.8.3.02>
- Nikiforova, A., Bicevskis, J., Bicevska, Z., & Oditis, I. (2020). User-Oriented Approach to Data Quality Evaluation. *JUCS - Journal of Universal Computer Science*, 26(1), 107–126. <https://doi.org/10.3897/jucs.2020.007>
- Nogueras-Iso, J., Lacasta, J., Urena-Camara, M. A., & Ariza-Lopez, F. J. (2021). Quality of Metadata in Open Data Portals. *IEEE Access*, 9, 60364–60382. <https://doi.org/10.1109/ACCESS.2021.3073455>
- Papadomichelaki, X., & Mentzas, G. (2012). e-GovQual: A multiple-item scale for assessing e-government service quality. *Government Information Quarterly*, 29(1), 98–109. <https://doi.org/10.1016/j.giq.2011.08.011>
- Raca, V., Velinov, G., Dzalev, S., & Kon-Popovska, M. (2022a). A Framework for Evaluation and Improvement of Open Government Data Quality: Application to the Western Balkans National Open Data Portals. *SAGE Open*, 12(2), 215824402211048. <https://doi.org/10.1177/21582440221104813>
- Raca, V., Velinov, G., Dzalev, S., & Kon-Popovska, M. (2022b). A Framework for Evaluation and Improvement of Open Government Data Quality: Application

- to the Western Balkans National Open Data Portals. *SAGE Open*, 12(2), 215824402211048. <https://doi.org/10.1177/21582440221104813>
- Redman, T. C. (2001). *Data Quality: The Field Guide*. Digital Press.
- Sanabria, M. A. O., Fernández, F. O. A., & Zabala, M. P. G. (2018a). Colombian Case Study for the Analysis of Open Data Government. *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, 389–394. <https://doi.org/10.1145/3209415.3209474>
- Sanabria, M. A. O., Fernández, F. O. A., & Zabala, M. P. G. (2018b). Colombian Case Study for the Analysis of Open Data Government. *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, 389–394. <https://doi.org/10.1145/3209415.3209474>
- Sebastian-Coleman, L. (2013). Measuring data quality for ongoing improvement: A data quality assessment framework. *Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework*, 1–324. <https://doi.org/10.1016/C2011-0-07321-0>
- Sidi, F., Shariat Panahy, P. H., Affendey, L. S., Jabar, M. A., Ibrahim, H., & Mustapha, A. (2012). Data quality: A survey of data quality dimensions. *2012 International Conference on Information Retrieval & Knowledge Management*, 300–304. <https://doi.org/10.1109/InfRKM.2012.6204995>
- Soylu, A., Corcho, Ó., Elvesæter, B., Badenes-Olmedo, C., Yedro-Martínez, F., Kovacic, M., Posinkovic, M., Medvešček, M., Makgill, I., Taggart, C., Simperl, E., Lech, T. C., & Roman, D. (2022). Data Quality Barriers for Transparency in Public Procurement. *Information*, 13(2), 99. <https://doi.org/10.3390/info13020099>
- Tepandi, J., Lauk, M., Linros, J., Raspel, P., Piho, G., Pappel, I., & Draheim, D. (2017). *The Data Quality Framework for the Estonian Public Sector and Its Evaluation* (pp. 1–26). https://doi.org/10.1007/978-3-662-56121-8_1
- Torchiano, M., Vetro, A., & Iuliano, F. (2017). Preserving the Benefits of Open Government Data by Measuring and Improving Their Quality: An Empirical

Study. *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, 144–153. <https://doi.org/10.1109/COMPSAC.2017.192>

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325–337. <https://doi.org/10.1016/j.giq.2016.02.001>

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86–95. <https://doi.org/10.1145/240455.240479>

Wang, R. Y., & Strong, D. M. (2015). Beyond Accuracy: What Data Quality Means to Data Consumers. <https://doi.org/10.1080/07421222.1996.11518099>, 12(4), 5–34. <https://doi.org/10.1080/07421222.1996.11518099>

Παράρτημα Α: Μετρικές Ανοιχτών Κυβερνητικών Δεδομένων

Πλαίσιο Ανοιχτών Κυβερνητικών Δεδομένων	Μετρικές
8 OGD Principles (2007)	completeness, Primary, Timely, Accessible, Machine processable, Non-Discriminatory (available without registration), Non-Proprietary (in terms of format) and License-free),
Sunlight Foundation (2010)	completeness, primacy, timeliness, ease of physical and electronic access, machine readability, nondiscrimination, use of commonly owned standards, licensing, permanence and usage costs.
Open Data Charter (2016)	Open by Default, Timely and Comprehensive, Accessible and usable, Comparable and Interoperable, For Improved Governance and Citizen Engagement, For Inclusive Development and Innovation
Tim Berners-Lee (2006)	5 stars open data

Παράρτημα Β: Μονάδες Μέτρησης Ποιοτικών Μετρικών

Unit of measure	Remark
%	Percentage
Number	Absolute number
Grade	Only the perception of people about the dimension can be measured. A grade can be number on a scale of 0-10 or 1-5.
Boolean	Yes/no or true/false
Duration	Expressed in seconds, minutes, hours, days, weeks or months.
Story	The value of the dimension cannot be expressed in a number and should be explained in a 'story'.

Source: <https://dama-nl.org/wp-content/uploads/2020/11/How-to-Select-the-Right-Dimensions-of-Data-Quality-v1.1-d.d.-14-Nov-2020.pdf>

Παράρτημα Γ: Πεδία ανά μετρική

➤ Completeness

Υποχρεωτικά πεδία CPSV-AP	Υποχρεωτικά πεδία API
Process ³⁶	
Public Service name	process_title
Public Service identifier	id
Public Service decription	description
Public Organization_preferredLabel	org_owner
	org_owner_private
Public Organization_spatial	provision_org_group

³⁶ Τα Bold-marked αφορούν τον κάθε πίνακα (οντότητα) που περιλαμβάνεται στο σχήμα (schema) του MITOS

	provision_org_owner_directory deadline_duration estimated_implementation_duration provision_org evidence_identification_type application_description
Process_steps	step_title step_description step_duration_min step_duration_max
Process_Evidences_cost	evidence_cost_calculator evidence_cost_code evidence_cost_description evidence_cost_payment_type
Process_Rules	rule_type rule_decision_number rule_decision_year rule_description rule_url
Process_Provision_Digital_Location	provision_digital_location_url

➤ **Availability**

Πεδία	Πίνακας
application_related_url	process
useful_link_URL	process_useful_links
rule_URL	process_rules

provision_digital_location_URL

digital_provisions

➤ **Consistency**

1	Αν ο αρμόδιος φορέας είναι ιδιωτικός (org_owner_is_private) δεν μπορεί να υπάρχει και Θεσμικός φορέας (org_owner)
2	Το πεδίο life_events πρέπει να είναι συμπληρωμένο
3	Αν έχω estimated_implementation_time, δεν μπορώ έχω και deadline_duration (ορισμένο από το νομικό πλαίσιο).
4	Όπου η Γλώσσα «Ελληνικά» είναι συμπληρωμένη δεν είναι σωστό βάσει Manual συντακτών
5	Το πεδίο provision_org_directory συμπληρώνεται εφόσον το provision_org είναι συμπληρωμένο
6	Το πρώτο βήμα μιας διαδικασίας δεν μπορεί να είναι βήμα διακλάδωσης (step_child)
7	Τα βήματα της διαδικασίας πρέπει να είναι min 3
8	Τα βήματα της διαδικασίας πρέπει να είναι min 3 (digital steps)
9	Όταν το πεδίο trigger είναι "Αιτούμενη" τότε trigger_type πρέπει να είναι συμπληρωμένο
10	Όταν το πεδίο rule_type δεν "Εγκύκλιος" τότε rule_ada πρέπει να είναι συμπληρωμένο
11	Το πεδίο validity_duration πρέπει να είναι συμπληρωμένο
12	Όταν το πεδίο step_num_id έχει την τιμή 1 τότε το step_exit δεν μπορεί να είναι 0
13	Το πρώτο ψηφιακό βήμα μιας διαδικασίας (step_digital_num_id είναι 1) δεν μπορεί να είναι βήμα διακλάδωσης
14	Όταν το πεδίο step_digital_num_id έχει την τιμή 1 τότε το step_digital_exit δεν μπορεί να είναι 0
15	Όταν το πεδίο evidence_num_id έχει τιμή 1 τότε το evidence_alternative πρέπει να είναι 0
16	Όταν το πεδίο evidence_alternative έχει τιμή 1 τότε το evidence_alternative_of πρέπει να είναι συμπληρωμένο
17	Όταν το πεδίο evidence_is_under_prerequisite έχει τιμή 1 τότε το evidence_prerequisite πρέπει να είναι συμπληρωμένο
18	Όταν το πεδίο evidence_cost_calculation είναι "Ποσοστό" τότε τα evidence_cost_min & evidence_cost_max) πρέπει να είναι συμπληρωμένα
19	Όταν υπάρχει κωδικός παραβόλου, θα πρέπει το evidence_cost_type να είναι e-Παράβολο
20	Όταν το πεδίο evidence_cost_type είναι "e-Παράβολο" τότε τα evidence_cost_calculator & evidence_cost_code πρέπει να είναι συμπληρωμένα
21	Όταν το πεδίο evidence_cost_type δεν είναι "e-Παράβολο" ή "e-Παράβολο κυμαινόμενο" τότε evidence_cost_calculation & evidence_cost_description & evidence_cost_payment_type & evidence_cost_min/max πρέπει να είναι συμπληρωμένα

22

Εντοπίζει τις διαδικασίες που έχουν first και last step σύμφωνα με συγκεκριμένα λεκτικά στο step_title