

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Δικτυακή Εφαρμογή - Πύλη Ανοικτών Δεδομένων
(crowdsourcing)



Του φοιτητή
Λούσι Άντζελο
Αρ. Μητρώου: 144347

Επιβλέπων
Χαράλαμπος Μπράτσας
Επίκουρος Καθηγητής

Ημερομηνία Ιανουάριος 2025

Τίτλος Π.Ε. Δικτυακή Εφαρμογή - Πύλη Ανοικτών Δεδομένων (crowdsourcing)

Κωδικός Π.Ε. 23109

Ονοματεπώνυμο φοιτητή Λούσι Άντζελο

Ονοματεπώνυμο εισηγητή Χαράλαμπος Μπράτσας

Ημερομηνία ανάληψης Π.Ε 06/03/2023

Ημερομηνία περάτωσης Π.Ε. 16/01/2025

Βεβαιώνω ότι είμαι ο συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή Λούσι Άντζελο που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

«Αφιέρωση»

Πρόλογος

Η παρούσα εργασία συντάχθηκε από τον φοιτητή Λούσι Άντζελο υπό την εποπτεία του κ. Μπράτσα Χαράλαμπου με τίτλο «Δικτυακή Εφαρμογή - Πύλη Ανοικτών Δεδομένων (crowdsourcing)». Βασικό θέμα το οποίο εξετάζεται στα πλαίσια της εργασίας αποτελούν οι Πύλες Ανοικτών Δεδομένων που είναι διαδικτυακές πλατφόρμες συλλογής δεδομένων του δημόσιου τομέα. Με την εκπόνηση της εργασίας δόθηκε η ευκαιρία για εμβάθυνση σε έναν ιδιαίτερα ενδιαφέρον τομέα, αυτό των ανοικτών δεδομένων καθώς και του Σημαιολογικού Ιστού. Κατά τη διάρκεια της έρευνάς, επιτυγχάνεται η απόκτηση γνώσεων σχετικά με τις τεχνολογίες και τα πρότυπα που διέπουν τις πύλες ανοικτών δεδομένων και τις εφαρμογές τους σε διάφορους τομείς, όπως η δημόσια διοίκηση. Τέλος ιδιαίτερη βαρύτητα αποτελεί η σημασία της ελεύθερης πρόσβασης σε πληροφορίες που αφορούν θέματα του δημόσιου βίου γεγονός που μπορεί να ενισχύσει δραστικά τη διαφάνεια, την καινοτομία και την οικονομική ανάπτυξη.

Περίληψη

Στις μέρες μας η συλλογή και εν συνέχεια η επεξεργασία των δεδομένων αποτελεί έναν τομέα εξελισσόμενο με ιδιαίτερο ενδιαφέρον και πολλαπλά οφέλη σε διάφορους τομείς της καθημερινότητας. Στα πλαίσια της παρούσας εργασίας στόχος είναι να μελετήσουμε εκτενώς τις διαδικτυακές πλατφόρμες οι οποίες ασχολούνται με την συλλογή δεδομένων που αφορούν τον δημόσιο τομέα. Με τη συμβολή διαφόρων πύλων ανοικτών δεδομένων γίνεται εφικτή η πρόσβαση σε σύνολα δεδομένων που συλλέγονται, παράγονται ή αποκτώνται από δημόσιους φορείς, επιτρέποντας τη χρήση τους για πολλαπλούς σκοπού. Πιο συγκεκριμένα ενισχύεται η διαφάνεια ενώ παράλληλα με την σωστή μελέτη των δεδομένων αυτών, καθιστάτε εφικτή η βελτίωση των δημόσιων υπηρεσιών και προάγεται η καινοτομία. Παράλληλα με τη συμβολή των διαφόρων τεχνολογιών σημασιολογικού ιστού (π.χ. RDF, OWL), η αξιολόγηση των δεδομένων είναι αποτελεσματικότερη, καθώς η αναζήτηση συγκεκριμένων πληροφοριών γίνεται ορθότερα, αλλά επιτυγχάνεται και η διασύνδεσή τους. Παράλληλα με τον όρο των συνδεδεμένων δεδομένων (Linked Data) γίνεται πιο εύκολη η διασύνδεση των δεδομένων που προέρχονται από πολλαπλές πηγές αυξάνοντας έτσι την χρησιμότητα τους δημιουργώντας συσχέτιση που οδηγούν σε καλύτερα αποτελέσματα. Στη συνέχεια εξετάζεται η χρήση των ανοικτών δεδομένων στη χώρα μας αλλά και στην Ευρωπαϊκή Ένωση, μέσω της ανάλυσης διάφορων ενεργειών που έχουν λάβει χώρα σε Ευρωπαϊκό επίπεδο, όπως η Οδηγία PSI, καθώς και τις εφαρμογές στην Ελλάδα μέσω πλατφορμών όπως το data.gov.gr και το σύστημα Διαύγεια. Επιπλέον, παρουσιάζεται η εγκατάσταση της πλατφόρμας CKAN, μιας λύσης διαχείρισης ανοικτών δεδομένων, η οποία διαθέτει δεδομένα από μετρήσεις ποιότητας νερού που έχουν πραγματοποιηθεί από την ΕΥΑΘ (Εταιρεία Ύδρευσης και Αποχέτευσης Θεσσαλονίκης) . Αυτές οι πλατφόρμες καθιστούν εφικτή στους χρήστες την ελεύθερη πρόσβαση σε δημόσια δεδομένα, ενισχύοντας την καινοτομία και τη διαφάνεια, ενώ παράλληλα συμβάλλουν στην οικονομική ανάπτυξη. Η ελεύθερη πρόσβαση σε δεδομένα του δημόσιου τομέα επιτρέπει σε πολίτες, ερευνητές και επιχειρήσεις να αξιοποιούν αυτά τα δεδομένα για τη βελτίωση της ποιότητας ζωής. Τέλος υπογραμμίζει ότι η ανάπτυξη και η σωστή διαχείριση των ανοικτών δεδομένων μπορεί να ενισχύσει την εμπιστοσύνη των πολιτών προς το κράτος, να ενισχύσει την οικονομική ανάπτυξη και να προωθήσει νέες μορφές συνεργασίας, όπως το crowdsourcing. Ωστόσο, υπάρχει η ανάγκη για συνεχείς προσπάθειες βελτίωσης της ποιότητας και της προσβασιμότητας των δεδομένων, καθώς και την υιοθέτηση διεθνών προτύπων για τη μεγιστοποίηση των οφελών.

Web Application - Open Data Portal (crowdsourcing)

LOUSI ANTZELO

Abstract

Nowadays, the collection and then the processing of data is an evolving field with great interest and multiple benefits in various areas of everyday life. In the context of this paper, the aim is to study extensively the online platforms that deal with the collection of data related to the public sector. With the contribution of various open data portals, it is possible to access datasets collected, produced or acquired by public entities, allowing their use for multiple purposes. In particular, it enhances transparency and, by properly studying this data, enables the improvement of public services and promotes innovation. At the same time, the contribution of various semantic web technologies (e.g. RDF, OWL) makes the evaluation of data more efficient, as the search for specific information is more accurate, but also achieves their interconnection. At the same time, the term Linked Data makes it easier to interconnect data from multiple sources thus increasing their usefulness by creating correlation leading to better results. In the following, the use of open data in Greece and in the European Union is examined through the analysis of various actions that have taken place at the European level, such as the PSI Directive, as well as the applications in Greece through platforms such as data.gov.gr and the Diavgeia system. In addition, the deployment of the CKAN platform, an open data management solution, which features data from water quality measurements carried out by EYATH (*Thessaloniki Water Supply & Sewerage Co S.A.*), is presented. These platforms enable users to freely access public data, enhancing innovation and transparency, while contributing to economic growth. Free access to public sector data allows citizens, researchers and businesses to use this data to improve the quality of life. Finally, it stresses that the development and proper management of open data can enhance citizens' trust in the state, boost economic growth and promote new forms of cooperation, such as crowdsourcing. However, there is a need for continuous efforts to improve the quality and accessibility of data, as well as the adoption of international standards to maximize the benefits.

Ευχαριστίες

Θα ήθελα να εκφράσω τις μου ευχαριστίες σε όλους όσους με στήριξαν κατά τη διάρκεια της συγγραφής της εργασίας. Αρχικά, ευχαριστώ τον επιβλέποντα καθηγητή μου κ. Μπράτσα Χαράλαμπου για την συμβολή του κατά τη διάρκεια της εκπόνησης της εργασίας καθώς η καθοδήγηση του και η πολύτιμη βοήθειά ήταν καθοριστικής σημασίας . Επίσης, ευχαριστώ την οικογένειά μου για τη συνεχή υποστήριξη και την ενθάρρυνσή τους καθ' όλη τη διάρκεια.

Περιεχόμενα

Πρόλογος.....	5
Περίληψη.....	6
Abstract	7
Ευχαριστίες	8
Περιεχόμενα	9
Κατάλογος Πινάκων.....	11
Συντομογραφίες.....	11
Κεφάλαιο 1ο: Ανοιχτά Δεδομένα	12
1.1 Εισαγωγή.....	12
1.2 Σκοπός, Στόχος και Δομή της εργασίας.....	13
1.3 Ανοιχτά Δεδομένα.....	13
1.3.1 Ορισμός.....	14
1.4 Βασικές Αρχές Ανοιχτών Δεδομένων	15
1.5 Είναι Σημαντικά Τα Ανοιχτά Δεδομένα.....	15
1.6 Πλεονεκτήματα των Ανοιχτών Δεδομένων.....	16
1.7 Ανοιχτά Δημόσια Δεδομένα.....	18
1.8 Άδειες χρήσης Ανοιχτών Δεδομένων.....	19
1.8.1 Creative Commons.....	19
1.8.2 Open Data Commons (ODC)	21
1.9 Επίλογος.....	21
Κεφάλαιο 2ο: Σημασιολογικός Ιστός και Ανοιχτά Συνδεδεμένα Δεδομένα(Linked Open Data) ..	23
2.1 ΕΙΣΑΓΩΓΗ	23
2.2 Σημασιολογικός Ιστός.....	24
2.3 Μεταδεδομένα.....	25
2.4 Οντολογίες.....	26
2.5 Ανοιχτά Δεδομένα και Σημασιολογικός Ιστός.....	28
2.6 Linked Data	28
2.7 Οφέλη Ανοιχτών Συνδεδεμένων Δεδομένων	29
2.8 5 Star Linked Data.....	30
Κεφάλαιο 3ο: Τα ανοιχτά Δεδομένα Στην Ελλάδα και στην ΕΕ	32
3.1 ΕΙΣΑΓΩΓΗ	32

3.2	ΟΔΗΓΙΑ PSI (PSI Directive (EU))	32
3.3	Ευρωπαϊκή Στρατηγική	33
3.4	Ανοιχτά Δεδομένα στην Ελλάδα	34
3.4.1	Data.gov.gr	36
3.4.2	Διαύγεια.....	37
3.4.3	GEODATA.....	38
3.5	Η Δράση του Open Knowledge Foundation	39
Κεφάλαιο 4ο:	Πύλες Ανοιχτών Δεδομένων	40
4.1	ΕΙΣΑΓΩΓΗ	40
4.1.1	API	41
4.1.2	DRUPAL	41
4.2	CKAN.....	41
4.2.1	Αρχιτεκτονική του CKAN.....	42
4.2.2	CKAN COMPONENTS	43
4.2.3	CKAN EXTENSIONS - PLUGINS	43
4.2.4	CKAN API	44
4.2.5	CKAN – DATASTORE - FILESTORE.....	44
4.3	Data Catalog Vocabulary (DCAT).....	46
4.3.1	DCAT-AP.....	47
4.3.2	CKAN – DCAT	48
4.4	Επίλογος.....	49
Κεφάλαιο 5ο:	Παραδείγματα με Δεδομένα της ΕΥΑΘ.....	51
5.1	ΕΙΣΑΓΩΓΗ	51
5.2	Γιατί τα Δεδομένα της ΕΥΑΘ ;.....	54
5.3	Λειτουργίες και API που χρησιμοποιήσαμε.....	54
5.3.1	Λειτουργίες Διαχειρίσεις Δεδομένων (BACK-END) API	55
5.3.2	Λειτουργία Εύκολης Αναζήτησης.....	60
5.3.3	DCAT-RDF - LINKED DATA (Παράδειγμα με τα δεδομένα της ΕΥΑΘ).....	61
5.3.4	Εργαλεία που Χρησιμοποιήθηκαν.....	63
5.4	Συμπεράσματα Εργασίας	64
5.5	Σχέδια Μελλοντικής Εξέλιξης του Συστήματος	65
BIBΛΙΟΓΡΑΦΙΑ.....		66

Κατάλογος Πινάκων

Πίνακας 1.1: Συχνότεροι συνδυασμοί Creative Commons.....	20
Πίνακας 1.2 : Οι κύριες άδειες του Open Data.....	21
Πίνακας 3.1: Συνολικές βαθμολογίες ωριμότητας συμμετέχουσες ΕΕ 2023 Europa.Eu	35

Συντομογραφίες

Δ.Ε.	Διπλωματική Εργασία
ΔΙΠΑΕ	Διεθνές Πανεπιστήμιο Ελλάδος
Π.Ε.	Πτυχιακή Εργασία

Κεφάλαιο 1ο: Ανοιχτά Δεδομένα

1.1 Εισαγωγή

Αρχικά, φανταστείτε πόσο πολύ πληροφορία παίρνουμε καθημερινά τριγύρω μας είτε από σχολεία ,δουλείες, φίλους, η ακόμα και από τον Παγκόσμιο Ιστό (WEB) η αλλιώς από το (Internet), η πληροφορία που μπορεί κάποιος να δεχθεί στην καθημερινότητα του είναι άπειρη και αυτήν την πληροφορία σαφώς μπορούμε να την ερμηνεύσουμε όπως εμείς ακριβώς θέλουμε και να την μοιραστούμε παράλληλα και με άλλους. Όμως για να έχουμε μια πληροφορία και ακόμα και στην πραγματική μας ζωή θα πρέπει να συνοδεύονται από δεδομένα, ένα τυχαίο εντελώς παράδειγμα για το πόσιμο νερό φτάνοντας σε σπίτι φίλου πχ. «Γιώργο το νερό στην περιοχή είναι πόσιμο μπορείς να πιείς από την βρύση χωρίς φόβο» έχουμε την πληροφορία ότι το νερό είναι όντως πόσιμο που την πληροφορία την πήραμε από κάποιον οργανισμό που έκανε της μετρήσεις « δεδομένα » και κάποιος ιδικός έβγαλε το πόρισμα ότι το νερό είναι πόσιμο. Αναλύουμε αυτό το παράδειγμα να κατανοήσουμε πόσο σημαντικά είναι τα δεδομένα αλλά όχι από μόνα τους δίνοντας τας αξία ,οπού η αξία στο παράδειγμά μας έδωσε ο ιδικός που έβγαλε το πόρισμά του νερού, και να τονίσουμε ότι η πληροφορία αποκτάτε όταν τα δεδομένα έχουν αξία, το πόρισμα στο παράδειγμά μας, όπου τα δεδομένα από μόνα τους δεν θα μπορούσαν να μας δώσουν την πληροφορία αυτή καθώς είναι ένα σύνολο δεδομένων ακατέργαστων αριθμών , συμβόλων ,διακριτών στοιχείων η ακόμα και λέξεων (φράσεων), καθώς αν δεν πάρουν αξία τα δεδομένα δεν μας είναι καθόλου πολύτιμά γενικότερα. Η πληροφορία από την άλλη δεν πάει μαζί με τα δεδομένα αφού δεν μπορεί να παραχθεί από μονή της.

Γενικότερα , η παροχή δεδομένων δίνοντας τα αξία ,παράγοντας έτσι την πληροφορία (πόρισμα), είναι αρκετά έως και πολύ σημαντική για της ανάγκες της καθημερινότητας μάς , την σωστή λήψη αποφάσεων για σκοπούς υγείας, διάφορες αναλύσεις για επιχειρήσεις και για πολλά αλλά πράγματα αφού η σωστή και η έγκυρη πληροφορία με την σωστή χρήση της αναδεικνύετε ότι μπορεί να προσφέρει πάρα πολλά πράγματα ωφέλημα στον άνθρωπο για διάφορους λογούς από υγείας μέχρι και την οικονομική ανάπτυξη της επιχείρησης η ακόμα και στην εξοικονόμηση του χρόνου μπορεί να επιτευχθεί.

Έτσι λοιπόν όπως αναφέραμε μερικούς λόγους καταλήγουμε ότι τα δεδομένα είναι σημαντικά ως προς το όφελος μας κυρίως όταν αυτά έχουν επίδραση από την ζωή μας όπως περιπτώσεις υγείας, η διάφορους νόμους , διακυβερνητικά θέματα ή γεωχωρικά δεδομένα όπου πρέπει να υπάρχει ελεύθερη πρόσβαση σε αυτά έτσι ώστε να μπορούν να χρησιμοποιηθούν από διάφορους οργανισμούς ,δημοσιογράφους η και πολίτες που επιθυμούν να έχουν πρόσβαση σε αυτά τα δεδομένα και να θεωρούνται «Ανοιχτά» . Σε αυτό το πλαίσιο, τα «Ανοιχτά Δεδομένα» είναι κρίσιμα, γιατί επιτρέπουν την ελεύθερη πρόσβαση σε πληροφορίες που μπορούν να βελτιώσουν τις αποφάσεις μας, να προωθήσουν την καινοτομία και να ενισχύσουν τη διαφάνεια. Είναι ένα θέμα προς συζήτηση καθώς τα με Ανοιχτά Δεδομένα η κοινωνία μπορεί να επωφεληθεί στο σύνολο της, διαχειρίζοντας καλύτερα τους πόρους και διάφορες προκλήσεις, λαμβάνοντας σωστότερες, γρηγορότερες και πιο αποτελεσματικές αποφάσεις. Στις παρακάτω παραγράφους της ενότητας θα αναφερθούμε για τα Ανοιχτά Δεδομένα για το πώς και γιατί είναι χρήσιμα , για τις άδειες χρήσης που πρέπει να ακολουθούνται τα δεδομένα, τις βασικές τους αρχές και για τον ορισμό του Ανοιχτού και για τα Δημόσια ανοιχτά Δεδομένα.

1.2 Σκοπός, Στόχος και Δομή της εργασίας

Ο κύριος σκοπός της εργασίας αυτής είναι η παρουσίαση και η κατανόηση της λειτουργίας της επιλεγμένης πύλης CKAN. Στόχος είναι να αναδειχθεί η χρησιμότητα και η σημαντικότητα των ανοιχτών δεδομένων και πώς αυτά μπορούν να συμβάλουν στην καθημερινή μας ζωή. Μέσω διαφορών παραδειγμάτων η εργασία επιδιώκει να αναλύσει την λειτουργικότητά του CKAN κατανοώντας αρχικά και την σημασιολογία των συνδεδεμένων δεδομένων, προκειμένου να καταδείξει την αποτελεσματικότητα των ανοιχτών δεδομένων στην προώθηση της διαφάνειας και της καινοτομίας.

Οι στόχοι της εργασίας αυτής ήρθαν σταδιακά με κάθε στόχο να βασίζετε και στον προηγούμενο στόχο και να δώσουν μια πιο ολοκληρωμένη και διάσταση στην κατανόηση του θέματος. Αρχικά, ο πρώτος στόχος της εργασίας είναι η εύκολη κατανόηση των ορισμό των Ανοιχτών Δεδομένων και για ποιον λόγο αυτά είναι σημαντικά ειδικά τα Δημοσία δεδομένα, και γιατί είναι σημαντικό να υπάρχουν αυτά τα δεδομένα στην μορφή που χρειάζεται για εύκολή πρόσβασή σε αυτών αλλά και το επιπλέον σημαντικό για τις άδειες χρήση που θα ακολουθούν τα δεδομένα αυτά. Βασιζόμενος στον προηγούμενο στόχο είναι σημαντικό για πιο τεχνικά θέματα οι έννοιες και η ανάλυση ως προς την εύκολη κατανόηση της λειτουργίας και το πώς τα ανοιχτά δεδομένα είναι θα πρέπει να είναι ακολουθώντας σημαντικά πρότυπα για την εύκολη αναζήτηση και οργάνωση αυτών σε συνδυασμό με τον Σημασιολογικό Ιστό και την απαραίτητη χρήση των συνδεδεμένων δεδομένων τηρώντας τα 5 επίπεδα του Συνδεδεμένων δεδομένων 5 Star Linked Data του Tim Berners-Lee. Ύστερα, ήρθε η ανάγκη να στην μελέτη των ανοιχτών δεδομένων στην Ελλάδα φτιάχνοντας την δομή την επίτευξη του επόμενου στόχου όπου είναι η ανάδειξη της πύλης CKAN αναλύοντας της αναλυτικά και την πραγματοποίηση στην δικιά μας πύλη ανοιχτών δεδομένων χρησιμοποιώντας παράδειγμα από δεδομένα της ΕΥΑΘ έτσι ώστε να προβάλλουμε για το πώς τα δεδομένα αυτά θα πρέπει να παρέχονται στο εύρη κοινό και το πώς αυτές συλλέχθηκαν και με τον τρόπο εισάγονται στην πύλη δημιουργώντας επίσης λόγου της ανάγκης που προέκυψε διάφορες μεθόδους για την αυτόματη εισαγωγή των νέων δεδομένων που δημοσιεύονται στην ΕΥΑΘ όπως και όπου αυτά χρειάζεται ανάλογα, εκτιμώντας με αυτών τον τρόπο τα οφέλη των ανοιχτών δεδομένων.

Συνοπτικά, η εργασία στοχεύει να παρέχει μια ολοκληρωμένη εικόνα για την έννοια των ανοιχτών δεδομένων και την σημασία τους επιδιώκοντας να αναδείξει τα οφέλη και τις προκλήσεις που συνδέονται με την υλοποίηση και χρήση των ανοιχτών δεδομένων, προσφέροντας παράλληλα πρακτικές προτάσεις για την ενίσχυση της αποτελεσματικότητάς τους.

1.3 Ανοιχτά Δεδομένα

Τα Ανοιχτά Δεδομένα αναφέρονται στις πληροφορίες που παρέχονται κατά κυρίως λόγο από τους δημόσιους φορείς για την πρόσβαση σε αυτά και την επαναχρησιμοποίηση τους από επιχρίσεις - επιχειρηματίες που μπορούν να τα χρησιμοποιήσουν αυτά τα δεδομένα είτε από προγραμματιστές για κυρίως λόγο την δημιουργία νέων τεχνολογικών υπηρεσιών για καινοτόμες πλατφόρμες ιδίως σε τομείς όπως είναι η υγειονομική περίθαλψη η εκπαίδευση η ακόμα και για τις μεταφορές δίνοντάς στις νέες τεχνολογίες αυτές προστιθεμένη αξία. Για να επιτευχθεί όμως κάτι τέτοιο θα πρέπει τα δεδομένα αυτά να είναι σωστά δημοσιευμένα και αναρτημένα στις πύλες των Ανοιχτών δεδομένων χωρίς κανέναν περιορισμό και προσβάσιμα σε όλους.

Η σωστή δημοσίευση των δεδομένων είναι πραγματικά ένας εξαιρετικός τρόπος για τη βελτίωση της διαφάνειας ενός οργανισμού και έπειτα με την παροχή της πληροφορίας σχετικά με την αξία των δεδομένων προς το ευρύ κοινό. Επιπλέον, έχουν κρίσιμο ρόλο στην έρευνα και την εκπαίδευση, παρέχοντας πρόσβαση σε τεράστιες ποσότητες πληροφοριών που μπορούν να χρησιμοποιηθούν για την

προώθηση της γνώσης η την μελέτη και την υποστήριξη της παγκόσμιας συνεργασίας. Συνοπτικά για την σωστή ανάρτηση των δεδομένων θα πρέπει τα δεδομένα να ακολουθάνε κάποιους κανόνες που έχουν ορισθεί έτσι ώστε να είναι προσβάσιμα σε όλους από όπου και να βρίσκονται χωρίς καμία δέσμευση τοποθεσίας η ακόμα και χρόνου , να είναι ελεύθερα ,να είναι προσβάσιμα και ευανάγνωστα από μηχανές για την επεξεργασία τους.

1.3.1 Ορισμός

Ο ορισμός «ανοικτά δεδομένα» αναφέρεται σε δεδομένα που είναι «ανοικτά» σε όλους, καθώς επιτρέπουν την χρήση τους και επαναχρησιμοποίησή τους από όλους. Ωστόσο, για να την σωστότερη χρήση αυτής της έννοιας «Ανοιχτά» η του όρου θα πρέπει να αναφερθεί ότι τα δεδομένα αυτά θα πρέπει να ακολουθάνε πρότυπα σύμφωνα με το Open Definition του Open Knowledge (OKN) (Ίδρυμα Ανοικτής Γνώσης) , όπως :

✚ Ανοικτά - Ελεύθερη Άδεια Χρήσης

Ο όρος «Ανοικτά» σημαίνει το γεγονός ότι τα δεδομένα δεν έχουν πνευματικά δικαιώματα ή πρέπει να υπάρχουν ελεύθερες άδειες για τη χρήση τους με τρόπους όπως η επαναχρησιμοποίηση, η κοινή χρήση, η τροποποίηση και η ανάμειξή τους με άλλα σύνολα δεδομένων να είναι εφικτή

✚ Προσβασιμότητα

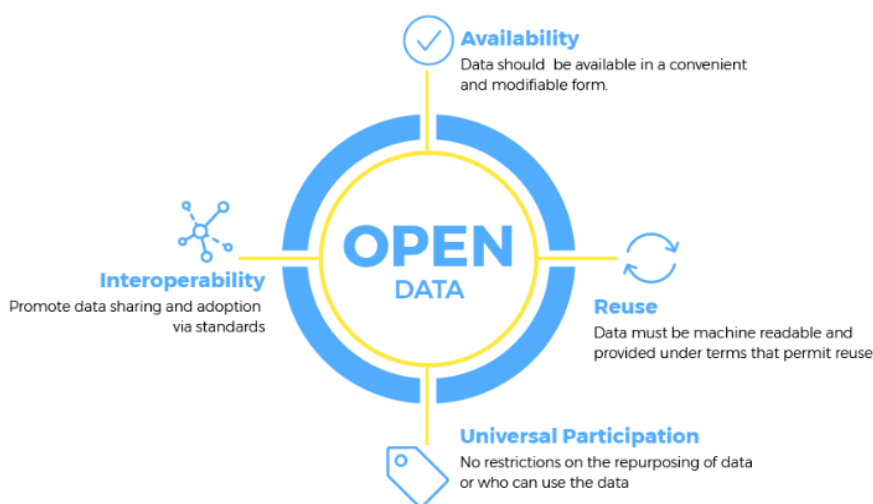
Διαθέσιμο σε όλους και δεν περιορίζεται σε συγκεκριμένες ομάδες, τοποθεσίες η οτιδήποτε άλλο.

✚ Machine Readability (‘Αναγνωσιμότητα’ από μηχανήματα)

Η μορφή θα πρέπει να είναι επεξεργάσιμη από μηχανήματα όπως υπολογιστές όπου τα στοιχεία του να μπορούν εύκολα να είναι προσβάσιμα ώστε να υπάρχει η δυνατότητα τροποποίησης και ανάγνωσης τους, για παράδειγμα να είναι σε τύπους δεδομένων (CSV,XML,JSON,RDF ,κ.λπ.)

✚ Ανοικτή μορφή (Ελευθέρω χωρίς κάποια χρέωση)

Ανοικτή μορφή είναι αυτή που δεν θέτει περιορισμούς, χρηματικούς ή άλλους, στη χρήση της και μπορεί να επεξεργαστεί πλήρως με τουλάχιστον ένα εργαλείο λογισμικού ανοικτού κώδικα.



Εικόνα 1.1 Πρότυπα Ανοιχτών Δεδομένων (opendata.info)

1.4 Βασικές Αρχές Ανοιχτών Δεδομένων

Τα Ανοιχτά Δεδομένα αποτελούν μια σημαντική εξέλιξη στην εποχή της πληροφορίας καθώς περνιέται η ιδέα και η νοοτροπία ότι τα δεδομένα θα πρέπει να είναι ελεύθερα και προσβάσιμα για όλους για την χρήση τους ή την επαναχρησιμοποίησή τους. Η έννοια αυτή δεν προωθεί μόνο την διαφάνεια αλλά και την καινοτομία που μπορεί να επιτευχθεί με βάση αυτών των δεδομένων αυτών. Για την σωστή λοιπών ανάρτηση αυτών των δεδομένων, τα δεδομένα αυτά θα πρέπει να ακολουθάνε κάποιες βασικές αρχές όπως,

- ❖ **Χρήση και Κόστους** τα δεδομένα αυτά θα πρέπει να είναι ανοιχτά αναρτημένες χωρίς να χρειάζεται κάποια εξειδικευμένη άδεια για την χρήση τους ή κάποια αγορά αδειας χρήσης,
- ❖ **Μακροπρόθεσμη διαθεσιμότητα** Τα σύνολα δεδομένων αυτών θα πρέπει να είναι διαχρονικά και να ενημερώνεται οι αλλαγές όταν αυτά αλλάζουν για οποιονδήποτε λόγο
- ❖ **Μη διάκριση** δεν θα πρέπει να υπάρχουν διακρίσεις ανάμεσα σε κάποιο πρόσωπο ή κάποια ομάδα όλοι την ίδια ελεύθερη πρόσβαση από όλους και από οπουδήποτε χωρίς καμία Διάκριση
- ❖ **Τροποποίηση** Θα πρέπει τα δεδομένα είναι ελεύθερα στην τροποποίηση τους χωρίς καμία επιπλέον άδεια για τα τροποποιημένα δεδομένα θα και να συμβάλλονται όπως και τα πρώτα αρχικά τους εφαρμογή για οποιονδήποτε ή αναδιανομή, η τροποποίηση και τη συνένωση των δεδομένων να μπορούν να γίνονται για οποιονδήποτε σκοπό δεν πρέπει να περιορίζετε κανέναν από το να χρησιμοποιήσει τα δεδομένα για συγκεκριμένο τομέα.
- ❖ **Διανομή** να επιτρέπετε τη διανομή των δεδομένων, είτε αυτό γίνεται για πώληση είτε δωρεάν, είτε διανέμεται μόνο του είτε ως μέρος μιας συλλογής που περιλαμβάνει δεδομένα από διαφορετικές πηγές.
- ❖ **Συλλογή** να επιτρέπει τη διανομή των δεδομένων με άλλα ξεχωριστά, χωρίς να επιβάλλονται περιορισμοί.

Ευανάγνωστα και Ευκολία Πρόσβασης τους από Μηχανήματα Τα δεδομένα θα πρέπει να αποθηκεύονται σε μορφές που επεξεργάζονται εύκολα από της μηχανές, να αποφεύγονται μορφές που δεν υποστηρίζουν εύκολη ανάγνωση από μηχανήματα, όπως τα PDF επίσης, θα πρέπει να συνοδεύονται από τεκμηρίωση που εξηγεί τη μορφή και τον τρόπο χρήσης των συνόλων αυτών για μαζική λήψη και πρόσβαση μέσω ενός API οπότε τα δεδομένα καθίστανται περισσότερο προσβάσιμα

Συνοψίζοντας λοιπόν, οι βασικές αρχές των ανοιχτών δεδομένων διασφαλίζουν την ελεύθερη χρήση, τροποποίηση και διανομή τους χωρίς περιορισμούς ή διακρίσεις. Αυτό μας επιτρέπει την ευρεία πρόσβαση και αξιοποίηση των δεδομένων προς όφελος της κοινωνίας και της καινοτομίας γενικότερα

1.5 Είναι Σημαντικά Τα Ανοιχτά Δεδομένα

«Τα ανοιχτά δεδομένα, ιδίως τα ανοιχτά κυβερνητικά, είναι ένας ανεκτίμητος πόρος πηγής πληροφοριών και δεδομένων που δεν έχουμε αξιοποιήσει ακόμα πλήρως γενικότερα» σύμφωνα με τον Οργανισμό Ανοιχτής Γνώσης Ελλάδος (Open Knowledge Greece). Αναρωτηθείτε λοιπόν, όταν κυβερνήσεις και διάφοροι οργανισμοί συλλέγουν ένα αρκετά μεγάλο όγκο δεδομένων για διάφορα θέματα όπως εργασίες που πρόκειται να γίνουν, αποφάσεις είτε διάφορες αναρτήσεις αυτά τα δεδομένα δεν θα έπρεπε να αναρτηθούν ως προς την κοινωνία έτσι ώστε να υπάρχει μια πλήρης διαφάνεια; Νομικά αυτά τα δεδομένα θα πρέπει να είναι ανοιχτά και προσβάσιμα σε όλους διότι αποτελούν δημοσιές πληροφορίες. Γιατί όμως αυτό είναι σημαντικό;

Η Σωστή δημοσιοποίηση των ανοιχτών δεδομένων στις πύλες έχει αποδεχθεί ότι είναι σημαντικό για αρκετούς τομείς που μπορεί να κάνουν την διαφορά είτε πρόκειται για κυβερνητικά είτε για επιχειρηματικά είτε ακόμα και οι πολίτες της χώρας μπορούν να επωφεληθούν εξίσου σημαντικά με τα Ανοιχτά Δεδομένα. Ένα παράδειγμά δημοσιεύοντας ένα σύνολο δεδομένων όσο αφορά την ποιότητα νερού σίγουρά υπάρχει μια **Διαφάνεια** ως προς το νερό και την πληροφορία που θα μας δώσουν τα δεδομένα αυτά και θα μπορούσαμε να αποφύγουμε τυχόν μη επιθυμητές καταστάσεις, **Ενίσχυση της συμμετοχής** οι πολίτες θα μπορούσαν να συμμετέχουν ενεργά στην καταγραφή δεδομένων για την ποιότητα του νερού, όπως πχ. προγράμματα παρακολούθησης των επιπέδων μόλυνσης σε τοπικά ποτάμια και λίμνες. **Βελτιωμένη απόδοση των δημόσιων υπηρεσιών** παίρνοντας υπόψιν το παράδειγμα μας αν τυχόν κάτι δεν συμφωνούσε με την ποιότητα του νερού ο δήμος της τοπικής περιοχής θα μπορούσε να πάρει γρηγορότερη, αποδοτικότερη και ασφαλέστερη πρωτοβουλία για την αντιμετώπιση του, αυτό επίσης συμβάλει ότι ίσως ένας οργανισμός θα μπορούσε να αναλάβει εν εντολή του τοπικού δήμου μια υπηρεσίας αντιμετώπιση τους χρησιμοποιώντας αυτά τα δεδομένα, σαφώς συμβάλουν στην **Βελτίωση των προϊόντων και υπηρεσιών** με αποτέλεσμα και ο κόσμος θα έχει **καλύτερη κατανόηση** ως προς την ποιότητα του νερού. Αυτός είναι σίγουρά ένας σημαντικός τρόπος ώστε η **Ποιότητα** του νερού να μένει σταθερή δείχνοντας ένα καλό πρόσωπο προς τας έξω κερδίζοντας την εμπιστοσύνη της κοινωνίας και σαφώς ενισχύοντας την **οικονομία** καθώς περισσότεροι άνθρωποι συμβάλουν σε μια τέτοια εργασία η ακόμα και λαμβάνοντας μέτρα σε μια πιθανή καταστροφή που μπορεί να επίδραση ακόμα και την οικονομία της χώρας και του τοπικού δήμου. Οι τομείς αυτοί δείχνουν πόσο ισχυρά μπορούν να είναι τα ανοικτά δεδομένα. Όταν περισσότερα δεδομένα γίνονται διαθέσιμα και περισσότεροι βρίσκουν δημιουργικούς και σωστούς τρόπους να τα χρησιμοποιήσουν, οι δυνατότητες είναι ατελείωτες.

Αυτή η επισκόπηση βασίζεται σε πληροφορίες από την Open Knowledge Greece, έναν οργανισμό που έχει δεσμευτεί να προωθή το άνοιγμα και τη διαφάνεια των δεδομένων σε διάφορους τομείς στην Ελλάδα.



Εικόνα 2 Γιατί Ανοιχτά Δεδομένα

1.6 Πλεονεκτήματα των Ανοιχτών Δεδομένων

Τα πλεονεκτήματα των ανοιχτών δεδομένων είναι πολυάριθμα κυρίως αναφέρονται στην αποτελεσματικότητα στις δημόσιες υπηρεσίες, την οικονομική ανάπτυξη στον ιδιωτικό τομέα ακόμα και στην συμβολή στην γενικότερη κοινωνική τάξη προσφέροντας ένα υψηλότερο επίπεδο υγείας, εκπαίδευσης, ασφάλειας και οικονομίας. Έτσι λοιπόν τα Ανοιχτά Δεδομένα μας προσφέρουν οφέλη όπως :

- ✚ *Ανοικτά, μόνιμα διαθέσιμα και σε οποιαδήποτε τοποθεσία χρήσης του διαδικτύου*
- ✚ *Κυβερνητικά, ο καθένας μπορεί να ενημερωθεί λεπτομερώς*
- ✚ *Αξιόπιστα, πιστοποιημένη η προέλευσή τους*
- ✚ *Ψηφιοποιημένα, μικρότερη αναμονή στην διαχείριση τους (αναζήτηση, χρήση, κ.α.)*
- ✚ *Αύξηση ποιότητας, αποτελεσματικότητας, διαφάνειας των δημόσιων υπηρεσιών*
- ✚ *Εξοικονόμηση κόστους, καθώς, για παράδειγμα, προβλέπεται ότι οι εθνικές κυβερνήσεις της ΕΕ+28 θα μπορούσαν να εξοικονομήσουν 1,7 δισεκατομμύρια ευρώ έως το 2020 (σύμφωνα με European Data Portal)*



Εικόνα 3 Πλεονεκτήματα Ανοιχτών Δεδομένων (European Data)

Η ανάπτυξη των Open Data έχει προσφέρει πολυάριθμα οφέλη στους δημοσίους οργανισμούς καθώς παρέχοντας οι φορείς τα δεδομένα προσφέρουν την μια πλήρη διαφάνεια, την καινοτομία όπως και την αποτελεσματική λήψη των αποφάσεων. Πιο συγκεκριμένα παρακάτω αναφέρονται τα πλεονεκτήματα των ανοιχτών δεδομένων σύμφωνα με το διάταγμα (Executive Order) της κυβέρνησης των ΗΠΑ το 2013

- Εξοικονόμηση χρόνου σε αιτήματα του νόμου περί ελευθερίας της πληροφόρησης
- Αποφυγή διπλής εσωτερικής έρευνας
- Δυνατότητα χρήσης δεδομένων που κατέχουν άλλοι οργανισμοί
- Βελτίωση για καλύτερες και πιο σίγουρες αποφάσεις με βάση τα δεδομένα
- Προσέλκυση προσοχής από το κοινό
- Δημιουργία εσόδων και νέων θέσεων εργασίας

Η δύναμη των ανοιχτών δεδομένων -όπως και κάθε άλλη χρήση δεδομένων- είναι η ικανότητά τους να οδηγούν και να υποστηρίζουν την βέλτιστη λήψη αποφάσεων, από άτομα έως οργανισμούς και ολόκληρες κοινότητες και έθνη. Το «άνοιγμα» των ανοιχτών δεδομένων μειώνει τις τριβές και αυξάνει την ικανότητα των ανθρώπων και των οργανισμών να χρησιμοποιούν αποτελεσματικά τα δεδομένα αυτά.

“ public repository intended to foster collaboration and promote the continual improvement of the Open Data Policy.” USA Government Executive Order May 9, 2013

«δημόσιο αποθετήριο που αποσκοπεί στην ενίσχυση της συνεργασίας και στην προώθηση της συνεχούς βελτίωσης της πολιτικής για τα ανοικτά δεδομένα.» Κυβέρνηση των ΗΠΑ διάταγμα Μάιος του 2013

1.7 Ανοικτά Δημόσια Δεδομένα

Τα ανοικτά δεδομένα τα οποία αναφέρονται στα δεδομένα που συλλέγονται, παράγονται είτε με την χρήση του αντίτιμου είτε δωρεάν και στην συνέχεια δημοσιεύονται για ελεύθερη χρήση, τα δεδομένα αυτά είναι γνωστά ως ανοικτά δημοσιάζ δεδομένα τα οποία καθορίζονται από αδειές χρήσης. Οι βασικές τους αξίες βασίζονται στις αρχές :

- **Ανοιχτού ορισμού**
- **Ελεύθερη πρόσβαση**
- **Ελεύθερης κυκλοφορίας των πληροφοριών**

Δεν είναι Λίγα τα οφέλη των ανοιχτών δεδομένων τα οποία και αναλύσαμε και παραπάνω πιο συγκεκριμένα έχοντας και χρησιμοποιώντας ανοικτά δημοσιάζ (κρατικά) δεδομένα οι **επιδόσεις μπορούν να αυξηθούν** σημαντικά στις δημοσιές υπηρεσίες χάρη στην διατομική ανταλλαγή δεδομένων. Φανταστείτε ότι δημοσιεύονται δεδομένα για την ποιότητα του νερού σε διαφορές περιοχές της Ελλάδας με μεταβλητές όπως παράδειγμά την θερμοκρασία του νερού τους δείκτες της βακτηρίαισής μόλυνσης η ακόμα και τα επίπεδα ρύπων, οι αρμόδιοι οργανισμοί θα μπορούν να παρακολουθούν συνεχώς την ποιότητα του νερού βελτιώνοντας έτσι με αυτό τον τρόπο την επίδοση του οργανισμού καθώς και κερδίζοντας την εμπιστοσύνη του κόσμου.

Επίσης, μεγάλο όφελος θα υπάρξει και στην οικονομία καθώς με την σωστή χρήση των ανοιχτών δημοσιών δεδομένων μόνο ανοδική θα μπορούσε να είναι η πορεία των καινοτόμων υπηρεσιών και των επιχειρηματικών μοντέλων. Ένα γρήγορο παράδειγμά με βάση το παραπάνω σε σχέση την ποιότητα του νερού, σκεφτείτε λοιπόν ότι με την δημοσίευση των δεδομένων αυτών μπορούμε και την διαχείριση των δεδομένων από τους Χ αρμοδίους θα μπορούσαμε να πούμε ότι η πρόληψη και διατήρηση των υδάτων σε φυσιολογικές τιμές ανάλογα με της μετρήσεις τους η διατήρησή της ποιότητας του νερού υψηλά, φανταστείτε τα χωράφια και την αγροτική παραγωγή η ποιότητα των λαχανικών και φρούτων δεν θα επηρεάζονται από το πιο σημαντικό τους συστατικό το νερό με αποτέλεσμα μεγαλύτερες παραγωγές.

Καθώς μπορούμε να πούμε ότι βελτιώνεται και η δυνατότητα του κράτους πρόνοιάς αυξάνοντας την εμπιστοσύνη της κοινωνίας που οφείλει να έχει διαφανέστερες πληροφορίες με αυξημένη προσβασιμότητα.



Εικόνα 4 Δεδομένα από το European Data

Τα ανοικτά δεδομένα σύμφωνα με το European Data (Επίσημος Ιστότοπος από την Ευρωπαϊκή Ένωση) ανεβάζουν σημαντικά την αποτελεσματικότητα της χρήσης των δεδομένων σε πραγματικό χρόνο κάνοντας πιο εύκολη την λήψη της απόφασης, όπως αναφέρετε από το European Data μπορούν να σώζουν 7000 ζωές τον χρόνο επιτυγχάνοντας την παροχή της ανάνηψης ενώ στους δρόμους μπορεί να μειωθεί σημαντικά ο χρόνος αναμονής στις 629 εκατομμύρια ώρες είναι ο χρόνος εξοικονόμησης αναμονής στους δρόμους της Ε.Ε (Ευρωπαϊκής Ένωσης) χρησιμοποιώντας δεδομένα σε πραγματικό χρόνο

1.8 Άδειες χρήσης Ανοιχτών Δεδομένων

Η κατάλληλη επιλογή αδειών χρήσης αποτελεί σημαντικό παράγοντα για την επίτευξη της ορθής διάδοσης και επαναχρησιμοποίησης δεδομένων για την επίτευξη των ανοικτών δεδομένων. Οι άδειες χρήσης καθορίζουν τους όρους υπό τους οποίους τα δεδομένα μπορούν να χρησιμοποιηθούν, τροποποιηθούν και διανεμηθούν, διασφαλίζοντας τόσο την ελευθερία πρόσβασης όσο και την προστασία των δικαιωμάτων των δημιουργών και παροχών των δεδομένων.

1.8.1 Creative Commons

Οι Άδειες Creative Commons (CC) είναι ευρέως γνώστες και χρησιμοποιούμενες για την παροχή των ελευθέρων αδειών χρήσης σε διάφορα είδη πνευματικής ιδιοκτησίας όπως κείμενα , εικόνες , βίντεο , μουσική , άρθρα και άλλα πολλά. Οι άδειες αυτές έχουν δημιουργηθεί από τον οργανισμό Creative Commons όπως και ονομάζονται, για την επιλογή μια σειρά επιλόγων που επιτρέπουν στους δημιουργούς των εκάστοτε έργων να καθορίζουν τους ορους όπου το ευρύ κοινό να μπορεί να χρησιμοποιεί την δουλειά του ιδιοκτήτη του έργου. Έτσι λοιπόν οι δημιουργοί μέσω αυτών των αδειών, οι δημιουργοί μπορούν να εξασφαλίσουν τον έλεγχο και την προστασία των έργων τους σε διάφορα στάδια, προωθώντας παράλληλα την κοινή χρήση και την επαναχρησιμοποίηση.

Τα Creative Commons πληρούνται από τα τέσσερα δικαιώματα (βασικοί όροι αδειών Creative Commons) που είναι τα παρακάτω :

Αναφορά (Attribution, BY)

Επιτρέπει στους χρήστες να μοιράζονται, να τροποποιούν και να χρησιμοποιούν το έργο, ακόμη και για εμπορικούς σκοπούς, υπό την προϋπόθεση ότι αναφέρουν τον αρχικό συγγραφέα.

Μη εμπορική χρήση (Non-commercial, NC)

Επιτρέπει τη χρήση και την τροποποίηση του έργου για μη εμπορικούς σκοπούς.

Παρόμοια διανομή (Share-Alike, SA)

Επιτρέπεται η αντιγραφή η διανομή και η τροποποίηση του έργου υπό την προϋπόθεση να έχει ακριβώς ίδιους όρους με το αρχικό έργο, εάν χρειαστούν άλλοι όροι πέρα του αρχικού έργου θα πρέπει να ζητηθεί η άδεια του δημιουργού.

Όχι παράγωγα έργα (No Derivatives, ND)

Επιτρέπεται η διανομή, αντιγραφή και η παρουσίαση μόνο του αρχικού έργου χωρίς τροποποίηση αυτού, αλλιώς θα χρειαστεί επικοινωνία με τον δημιουργό.

Οι συχνότεροι συνδυασμοί δικαιωμάτων που χρησιμοποιούνται γενικότερα στα έργα είναι :

ΣΥΝΤΟΜΟΓΡΑΦΙΑ	Εικόνα	ΠΕΡΙΓΡΑΦΗ
CC BY		Κατάλληλο για έργα που θέλουν να χρησιμοποιηθούν και να διατεθούν ευρέως, ακόμη και εμπορικά, με την προϋπόθεση της αναφοράς του δημιουργού.
CC BY-SA		Ιδανικό για έργα που υποστηρίζουν την ανοικτή συνεργασία και την ελεύθερη αντιγραφή υπό τους ίδιους όρους, κάνοντας αναφορά τον δημιουργό
CC BY-NC		Κατάλληλο για έργα που επιθυμούν να επιτραπεί η χρήση τους μόνο για μη εμπορικούς σκοπούς, αναφέροντας τον δημιουργό
CC BY-NC-SA		Είναι Κατάλληλο για έργα που δεν επιτρέπουν την εμπορική χρήση αλλά επιτρέπουν την χρήση τους και την τροποποίηση υπό την προϋπόθεση να πληρούν ίδιους όρους με το αρχικό έργο, αναφέροντας τον δημιουργό
CC BY-ND		Κατάλληλο για έργα που θέλουν να επιτρέψουν μόνο τη διάδοσή τους χωρίς τροποποίηση, κάνοντας αναφορά και εδώ τον δημιουργό.
CC BY-NC-ND		Χρησιμοποιείτε σε έργα που θέλουν να επιτρέψουν την διανομή τους χωρίς τροποποίηση και για μη εμπορικούς σκοπούς αναφέροντας τον δημιουργό του έργου

Αν και οι άδειες Creative Commons CC έχουν σχεδιαστεί κατά κύριο λόγο για έργα δημιουργικά, χρησιμοποιούνται και για τα δεδομένα. Ωστόσο, δεν είναι σαφείς για τις βάσεις δεδομένων και μπορεί να δημιουργήσουν νομική ασάφεια όταν εφαρμόζονται σε δεδομένα, καθώς τα δεδομένα δεν θεωρούνται δημιουργικό έργο σύμφωνα με τα πνευματικά δικαιώματα. Αυτό μπορεί να περιορίσει την

επαναχρησιμοποίηση και τη διανομή δεδομένων με σαφήνεια και προστασία, για αυτό τον λόγο δημιουργήθηκαν τα **Open Data Commons (ODC)**

1.8.2 Open Data Commons (ODC)

Τα **Open Data Commons (ODC)** είναι μια σειρά από άδειες χρήσης ειδικά σχεδιασμένες για να διευκολύνεται η κοινή χρήση, η επαναχρησιμοποίηση και η επαναδιάδοση των δεδομένων. Οι άδειες ODC αναπτύχθηκαν από το Open Knowledge Foundation (OKNF) για να αντιμετωπίσουν τις νομικές προκλήσεις που σχετίζονται με τις βάσεις δεδομένων και τα δομημένα δεδομένα από τις άδειες Creative Commons CC, έτσι ώστε να παρέχουν ένα σαφές και τυποποιημένο πλαίσιο που εξασφαλίζει την ελεύθερη πρόσβαση στα δεδομένα και τη χρήση τους, διασφαλίζοντας παράλληλα τα δικαιώματα των δημιουργών των δεδομένων. Οι άδειες αυτές ανεξάρτητος τύπου είναι σχεδιασμένα για την διασφάλιση της ανοχής της διάδοσης, δημιουργίας και επαναχρησιμοποίηση ή τροποποίηση διαφόρων έργων χρησιμοποιώντας της βάσης δεδομένων δεδομένων.

Οι κύριες άδειες του Open Data Commons είναι :

Τύπος Άδειας (ODC)	ΕΠΕΞΗΓΗΣΗ ΤΟΥ
1) Open Data Commons Attribution License	Παρόμοια με την CC BY, επιτρέπεται η κοινή χρήση, επαναχρησιμοποίηση, δημιουργία, τροποποίηση των δεδομένων με την μόνη απαίτηση την αναφορά του παρόχου των δεδομένων
2) Open Data Commons Public Domain Dedication and License (PDDL)	‘Απελευθερώνει’ τα όλα δεδομένα, επιτρέποντας την κάθε χρήση τους για οποιονδήποτε σκοπό χωρίς περιορισμούς
3) Open Data Commons Open Database License (ODbL)	Επιτρέπει στους χρήστες την κοινή χρήση, τροποποίηση και την χρησιμοποίηση των δεδομένων αρκεί να μένουν υπό τους ίδιους όρους. (Share – A like).

Η ανάπτυξη του Open Data Commons έχει παίξει καθοριστικό ρόλο στα ανοιχτά δεδομένα καθώς έχει λυθεί το νομικό κομμάτι των δεδομένων αυτών ενθαρρύνοντάς με αυτόν τον τρόπο την κοινή χρήση των δεδομένων, η βελτίωση της προσβασιμότητας στα δεδομένα αυτά είναι πιο φιλικά και όλα αυτά διασφαλίζοντας την ακεραιότητα των δεδομένων αυτών. Η συνεχής ανάπτυξη και διατήρηση των αδειών από το Open Knowledge Foundation επωφελή σημαντικά όλο ένα και περισσότερο τα ανοιχτά δεδομένα.

1.9 Επίλογος

Η συλλογή δεδομένων αποτελεί μια διαδικασία ιδιαίτερα εύκολη, καθώς ο όγκος των δεδομένων είναι μεγάλος και οι πληροφορίες εισέρχονται από παντού καθημερινά. Όπως καταγράφεται και στο παραπάνω κείμενο η συλλογή των δεδομένων αποκτά μεγαλύτερη σημασία όταν διατίθεται στο κοινό προκειμένου να έχει τη δυνατότητα να τις αξιοποιήσει. Επομένως τα ανοιχτά δεδομένα προσφέρουν μια σειρά δυνατοτήτων τόσο στην κοινωνία και όσο και στην οικονομία. Στο σημείο αυτό να επισημάνουμε ότι βασική προϋπόθεση είναι τα δεδομένα να δημοσιεύονται με τρόπο εύκολα αναγνώσιμο από μηχανές, ώστε να ενισχύεται η διαφάνεια και η καινοτομία. Παράλληλα

συμπεραίνουμε ότι όταν τα δεδομένα διατίθενται δημοσίως με τον σωστό τρόπο και ακολουθώντας όσο το δυνατόν καλύτερα τους κανόνες σωστής αποθήκευσης μπορούμε να έχουμε πολλαπλά οφέλη, όπως η βελτίωση της λήψης αποφάσεων, η δημιουργία νέων τεχνολογικών εφαρμογών και η ενίσχυση της εμπιστοσύνης των πολιτών προς το κράτος. Για το λόγο αυτό αποτελεί βασική προϋπόθεση να τηρούνται οι βασικές αρχές που διέπουν την αποθήκευση ανοιχτών δεδομένων, όπως η ελεύθερη πρόσβαση, η μακροπρόθεσμη διαθεσιμότητα και η δυνατότητα διανομής των δεδομένων χωρίς περιορισμούς. Επίσης είναι σημαντικό να τονιστεί η συμβολή των ανοιχτών δεδομένων σε έναν ακόμα βασικό τομέα που επηρεάζει τη δημόσια ζωή, αυτόν της οικονομικής ανάπτυξης. Η συμβολή των δεδομένων αυτών θα πρέπει να αποσκοπεί στην ενίσχυση των δημόσιων υπηρεσιών και στην προώθηση καινοτόμων υπηρεσιών, όπως στην υγειονομική περίθαλψη, την εκπαίδευση, την παροχή δημοσίων έργων σε οδικά δίκτυα και άλλα. Κάτι τέτοιο ενισχύεται και με τη συλλογή δεδομένα σε πραγματικό χρόνο, συμβάλλοντας δραστικά στη βελτίωση της ποιότητας της ζωής των πολιτών, βοηθώντας να μην σπαταλούνται διάφοροι πόροι αλλά και πολύτιμος χρόνος. Τέλος τα ανοιχτά δεδομένα αποτελούν ανεκτίμητο πόρο για την κοινωνία, προσφέροντας οφέλη σε όλους τους τομείς και βελτιώνοντας τη λήψη αποφάσεων, την καινοτομία και την οικονομική ανάπτυξη. Εν κατακλείδι η σημασία των ανοιχτών δεδομένων, καλύπτει ένα ευρύ φάσμα από την υποστήριξη της διαφάνειας μέχρι την βελτίωση της κοινωνικής ζωής σε πολλαπλά επίπεδα.

Κεφάλαιο 2ο: Σημασιολογικός Ιστός και Ανοιχτά Συνδεδεμένα Δεδομένα (Linked Open Data)

2.1 ΕΙΣΑΓΩΓΗ

Ο τρόπος με τον οποίο οι άνθρωποι βρίσκουν πράγματα στο διαδίκτυο εξελίσσεται από την απλή αναζήτηση ελεύθερου κειμένου και λέξεων κλειδιών σε πιο εξελιγμένες σημασιολογικές τεχνικές τόσο για την αναζήτηση όσο και για την πλοήγηση. Ένα παράδειγμα αυτής της εξέλιξης των μηχανών αναζήτησης, όπως η **Google**. Στην αρχή, οι αναζητήσεις βασίζονταν σε απλές λέξεις-κλειδιά, όπου το απλά έψαχνε τις συγκεκριμένες λέξεις μέσα στο περιεχόμενο των ιστοσελίδων. Σήμερα οι αναζητήσεις έχουν γίνει πιο σημασιολογικές, με τα συστήματα να κατανοούν το νόημα πίσω από τις λέξεις. Για παράδειγμα, αν κάποιος αναζητήσει "Βαθμολογία Greek Superliga", η μηχανή αναζήτησης δεν θα αναζητά απλά τις λέξεις "Βαθμολογία" και "Greek" "Superliga" στις ιστοσελίδες, αλλά θα κατανοήσει ότι ο χρήστης θέλει πληροφορίες για την βαθμολογία του Ελληνικού πρωταθλήματος και δεν θα του φέρει αποτελέσματα από ισότοπους που απλά περιέχουν την βαθμολογία μέσα. Η μετάβαση αυτή από την απλή αναζήτηση κειμένου στη σημασιολογική κατανόηση επιτρέπει μεγαλύτερη ακρίβεια και χρήσιμα αποτελέσματα, κάνοντας την πλοήγηση στο διαδίκτυο πιο εύκολη και αποδοτική για τους χρήστες.

Ο Σημασιολογικός Ιστός, που εισήχθη από τον Tim Berners-Lee τον άνθρωπο που είχε την ιδέα του Παγκόσμιου Ιστού (World Wide Web), οραματίστηκε πλέον ένα μέλλον όπου τα δεδομένα θα είναι εμπλουτισμένα με νόημα, επιτρέποντας στις μηχανές να κατανοούν και να επεξεργάζονται τις πληροφορίες με τρόπο που να είναι πιο ουσιαστικός και χρήσιμος για τους χρήστες.

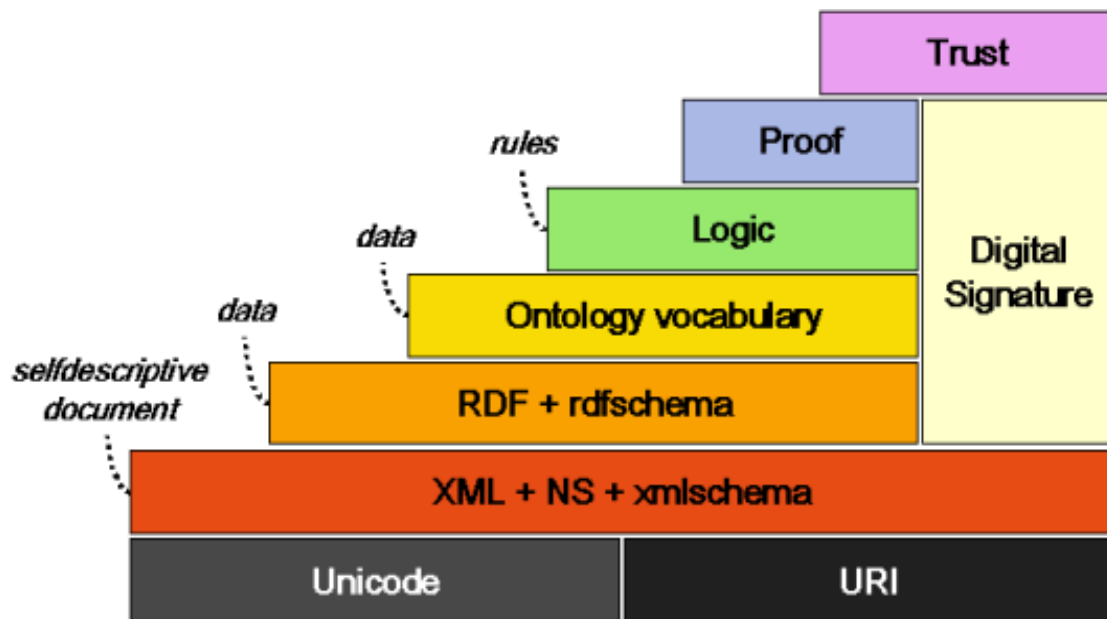
Ο Σημασιολογικός Ιστός περιγράφεται συχνά ως ένας ιστός δεδομένων και όχι ως ένας ιστός εγγράφων, προσθέτοντας επίπεδα νοήματος στα δεδομένα, επιτρέποντας στις μηχανές να τα κατανοούν και να τα χειρίζονται με τέτοιο τρόπο ίσο με αυτό του ανθρώπου. Η ιδέα του σημασιολογικού ιστού ήταν να ορίζονται τα δεδομένα και να μπορούν να χρησιμοποιηθούν για διάφορους αυτοματισμούς και για την επαναχρησιμοποίηση των δεδομένων. Αυτό είναι ιδιαίτερα σημαντικό στο πλαίσιο των ανοικτών δεδομένων, όπου μεγάλα σύνολα δεδομένων διατίθενται στο κοινό για χρήση από οποιονδήποτε. Τα συνδεδεμένα δεδομένα (Linked Data), ένα υποσύνολο των ανοικτών δεδομένων, ενισχύουν περαιτέρω το γεγονός αυτό, διασφαλίζοντας ότι αυτά τα σύνολα δεδομένων δεν είναι μόνο προσβάσιμα αλλά και διασυνδεδεμένα μεταξύ τους δημιουργώντας έναν πλούσιο και περιηγήσιμο ιστό δεδομένων.

Σε αυτήν την ενότητα θα εστιάσουμε στον Σημασιολογικό Ιστό, περιγράφοντας το RDF (Resource Description Framework), η XML (Extensible Markup Language), και το OWL (Web Ontology Language). Στη συνέχεια θα συζητήσει η σχέση του Σημασιολογικού Ιστού με τα Ανοιχτά Δεδομένα, τονίζοντας τα οφέλη από την εφαρμογή των αρχών του Σημασιολογικού Ιστού σε αυτές των Ανοιχτών Δεδομένων καθώς και η έννοια των Ανοιχτών Συνδεδεμένων Δεδομένων βλέποντας (Linked Data).

2.2 Σημασιολογικός Ιστός

Ο Σημασιολογικός Ιστός είναι μπορούμε να πούμε ότι είναι μια επέκταση του απλού παγκοσμίου ιστού (WWW), όπου φέρνει τη δομή και ουσιαστικό περιεχόμενο στα δεδομένα αλλά ακόμα και στις ιστοσελίδες. Η αρχική ιδέα και η λογική αυτής της τεχνολογίας είναι να περιέχονται **μεταδεδομένα** στις πληροφορίες που έχουν δημοσιευτεί, να είναι κοινά για όλους αλλά και να μπορούν να αναγνωρίζονται από της μηχανές για καλύτερη επεξεργασία ακόμα και για καλύτερη συλλογή των δεδομένων τους και την ανταπόκριση σε πολύπλοκα αιτήματα.

Η ανάπτυξη του Σημασιολογικού Ιστού προχωρά σε μικρά βήματα ως προς το τελικό, το οποίο βήμα χτίζει και ένα επίπεδο και πηγαίνει σε άλλο επίπεδο αυτό μια πολλυεπίπεδη προσέγγιση η μια διαστρωματωμένη προσέγγιση όπου αναφέρεται συχνά και αυτό επειδή, είναι ευκολότερο υπάρξει μία ομοφωνία χρησιμοποιώντας περισσότερά μικρά βήματα παρά μεγάλα.



Εικόνα 5 Επίπεδα Σημασιολογικού Ιστού

Το καθένα από τα οποία εξυπηρετεί έναν ξεχωριστό σκοπό στην οικοδόμηση ενός πιο ευφυσούς και διασυνδεδεμένου ιστού. Ακολουθεί μια ανάλυση αυτών των επιπέδων:

1. **Unicode και URI:** Αυτό το επίπεδο εξασφαλίζει τη χρήση διεθνών συνόλων χαρακτήρων και παρέχει μοναδικά αναγνωριστικά (URI) για τα αντικείμενα του Σημασιολογικού Ιστού. Τα URI είναι απαραίτητα για την αναφορά και τη διάκριση των πόρων.
2. **XML:** Η XML, μαζί με τα namespaces και τους ορισμούς σχημάτων, διευκολύνει την ενσωμάτωση των προτύπων του Σημασιολογικού Ιστού με άλλες τεχνολογίες που βασίζονται στην XML. Αυτό το επίπεδο παρέχει τη δομή και τη μορφή που απαιτούνται για το χειρισμό δεδομένων με συνεπή τρόπο.
3. **RDF και RDF Schema:** Σε αυτό το επίπεδο, το RDF (Resource Description Framework) επιτρέπει τη δημιουργία δηλώσεων σχετικά με πόρους που προσδιορίζονται από URIs. Το RDF Schema (RDFS) βοηθά στον καθορισμό των λεξιλογίων και των σχέσεων μεταξύ διαφορετικών τύπων πόρων, επιτρέποντας πλουσιότερες περιγραφές και συνδέσεις δεδομένων.

4. **Οντολογία:** Υποστηρίζει την ανάπτυξη και την εξέλιξη των λεξιλογίων, ορίζοντας σύνθετες σχέσεις μεταξύ διαφόρων εννοιών.
5. **Κανόνες:** Ορίζει λογικούς κανόνες για την εξαγωγή νέων πληροφοριών από τα υπάρχοντα δεδομένα
6. **Επίπεδο ψηφιακής υπογραφής:** Αυτό το επίπεδο είναι υπεύθυνο για την επαλήθευση της ακεραιότητας των εγγράφων και την ανίχνευση τυχόν αλλαγών. Εξασφαλίζει αξιόπιστα και αμετάβλητα κατά τη μετάδοση.

Οι πράκτορες πρέπει πλήρη γνώση ενός επιπέδου και να μπορούν να αξιοποιούν πλήρως τις πληροφορίες των από κάτω επιπέδων, καθώς και να έχουν την δυνατότητα της χρήσης μέρους πληροφοριών από τα πιο πάνω επίπεδα

Γενικότερα, ο σημασιολογικός ιστός παρέχει τεχνολογίες και πρότυπα έτσι ώστε να γινεί εφικτό η προσθήκη της πληροφορίας στα δεδομένα που είναι ήδη στον υπάρχοντά ιστό, με σκοπό την αξιοποίηση των δεδομένων, όπως :

- **Μεταδομένα , αναγνώριση-εξαγωγή πληροφορίας**
- **Οντολογίες – αποτελεσματικότερη αναζήτηση**
- **Λογική, Συλλογιστική και κανόνες- επεξεργασία πληροφορίας και εξαγωγή συμπεράσματος**
- **Πράκτορες – Απαντήσεις σε χρήστες χρησιμοποιώντας τα παραπάνω.**

2.3 Μεταδεδομένα

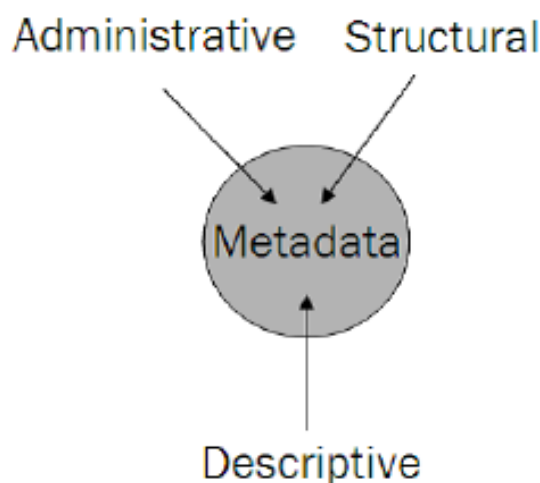
Όπως αναφέρθηκε παραπάνω τα μεταδεδομένα αποτελούν ένα βασικό πρότυπο που παρέχεται από τον Σημασιολογικό Ιστό. Η κατηγορία αυτή αποτελείται από δεδομένα που περιγράφουν το πώς και πότε δημιουργήθηκαν οι πληροφορίες, ποιος τις δημιούργησε και ποια πρότυπα χρησιμοποιήθηκαν. Αναλυτικότερα τα μεταδεδομένα λειτουργούν ως πληροφορίες οι οποίες περιγράφουν άλλα δεδομένα, βοηθώντας στην καλύτερη διαχείριση και ανάκτηση των δεδομένων. Στο σημείο αυτό να αναφέρουμε ότι σημερινοί οι περιορισμοί της HTML δυσκολεύουν τις μηχανές να κατανοήσουν πλήρως το περιεχόμενο των ιστοσελίδων. Με τη χρήση καλύτερων γλωσσών και μεταδεδομένων , το περιεχόμενο του ιστού μπορεί να γίνει ευκολότερα κατανοητό και να διαμοιραστεί. Παρακάτω αναπτύσσεται ένα παράδειγμα μεταδεδομένων σχετικά με ένα δημόσιο φορέα, με σκοπό να γίνει κατανοητός ο όρος. Ένα σύνολο πληροφοριών θα μπορούσε να αφορά τα δεδομένα που αποτελούν στο Υπουργείο Τουρισμού και αποθηκεύονται σε διάφορες μορφές. Αναλυτικότερα τα μεταδεδομένα αυτά θα μπορούσαν να μας δώσουν πληροφορίες σχετικά με τα χρονικά διαστήματα που ενημερώθηκε η βάση δεδομένων του Υπουργείου π.χ. έτος . Επίσης θα μπορούσε να περιλαμβάνει διάφορα δημογραφικά στοιχεία των νομών, όπως για παράδειγμα το σύνολο των τουριστών ανά νόμο που επισκέφθηκαν την Ελλάδα τον μέσο όρο της ηλικίας ,ανά διάστημα μηνών (low Season,Medium Season , High Season) . Επομένως βοηθούν στην εξόρυξη διαφόρων συμπερασμάτων, διευκολύνοντας την εύρεση και τη χρήση τους σε πολλαπλά συστήματα ή πλατφόρμες που διαχειρίζονται αυτά τα δεδομένα. Άρα στο συγκεκριμένο παράδειγμα οι πληροφορίες που λαμβάνουμε παρέχουν δεδομένα για τον αριθμό τουριστών ανά συγκεκριμένο χρονικό διάστημα και νόμο, και την κατανομή των αυτών ανάλογα την ηλικία.

Για την καλύτερη κατανόηση του όρου να αναφέρουμε ότι τα μεταδεδομένα μπορούν να διακριθούν σε τρεις βασικές κατηγορίες. Με βάση τις ανάγκες των χρηστών κάθε κατηγορία παρέχει διαφορετικές

δυνατότητες. Με αυτόν τον τρόπο δυναμικά προσφέρετε η διαχείριση των δεδομένων, κάνοντας τα πιο εύχρηστα και αποδοτικά τόσο για τους χρήστες όσο και για τα συστήματα που τα επεξεργάζονται.

Ακολουθούν οι τρεις βασικές κατηγορίες:

1. **Δομικά Μεταδεδομένα (Structural Metadata):** Η κατηγορία αυτή περιγράφει τον τρόπο με τον οποίο γίνεται η οργάνωση των δεδομένων αλλά και των πως είναι δομημένα, όπως για παράδειγμα η δομή μιας βάσης δεδομένων.
2. **Περιγραφικά Μεταδεδομένα (Descriptive Metadata):** Μια ακόμα σημαντική κατηγορία η οποία επεξηγεί τα δεδομένα. Σκοπός τους είναι να γίνεται πιο εύκολη η αναγνώριση τους ώστε να διευκολύνεται η αναζήτηση με τη χρήση λέξεων κλειδιών
3. **Διαχειριστικά Μεταδεδομένα (Administrative Metadata):** Τέλος έχουμε τα διαχειριστικά μεταδεδομένα τα οποία συμβάλουν στη διαχείριση αλλά και τη συντήρηση των δεδομένων.



Εικόνα 2.1 Μεταδεδομένα

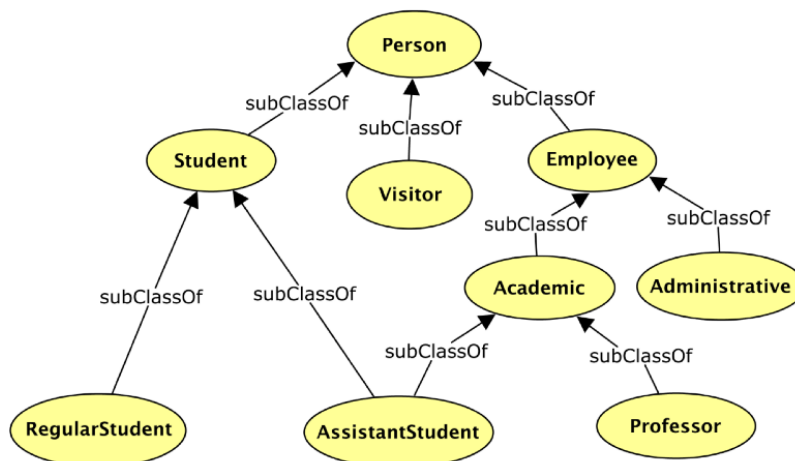
2.4 Οντολογίες

Η οντολογία παρέχει ένα δομημένο πλαίσιο για την περιγραφή της γνώσης σε έναν τομέα, χρησιμοποιώντας ένα πεπερασμένο σύνολο λιστών και τις σημασιολογικές τους σχέσεις. Μία Οντολογία αποτελείται από στοιχεία όπως έννοιες (κλάσεις), σχέσεις (ιεραρχικές κλάσεις). Οι οντολογίες συμβάλλουν στην τυποποίηση της αναπαράστασης της γνώσης, διευκολύνοντας την κοινή κατανόηση μεταξύ διαφορετικών συστημάτων και επίσης, είναι χρήσιμη για την βελτίωση της αναζήτησης στον ιστό, οι μηχανές αναζήτησης μπορούν να αναζητήσουν στοιχεία που αναφέρονται σε μια ακριβή έννοια κάποιας οντολογίας αντί να συλλέξουν όλες τις σελίδες που αναφέρονται.

Οι σημαντικότερες γλώσσες οντολογιών για τον Ιστό είναι οι εξής

- I. Το **RDF (Resource Description Framework)** είναι ένα πλαίσιο για την αναπαράσταση πληροφοριών σχετικά με πόρους στον ιστό. Παρέχει απλή σημασιολογία και τα δεδομένα μπορούν να παρασταθούν με XML

- II. **To RDF Schema (RDFS)** επεκτείνει το RDF παρέχοντας ένα βασικό σύστημα τύπων για τα δεδομένα RDF, ορίζοντας κλάσεις και ιδιότητες και προσδιορίζοντας ιεραρχικές σχέσεις.
- III. **Η OWL (Web Ontology Language)** βασίζεται στις RDF και RDFS για να προσφέρει πλουσιότερες σημασιολογικές δομές. Είναι μια πλουσιότερη γλώσσα περιγραφής των ιδιοτήτων των κλάσεων όπως τις σχέσεις μεταξύ τους ή την **Πληθικότητα** (cardinality) η ακόμα και την πλουσιότερη τυποποίηση των ιδιοτήτων ακόμα και τα χαρακτηριστικά αυτών , επιτρέποντας πιο διαφοροποιημένη και εξελιγμένη αναπαράσταση και εξαγωγή συμπερασμάτων γνώσης στον ιστό.



Εικόνα 2.2 Μία ιεραρχία με τις έννοιες

2.4.1.1 Λογική

Ένας βασικός παράγοντας που καθιστά εφικτή την εξαγωγή συμπερασμάτων από δεδομένα αποτελεί η λογική. Η λογική αναπτύχθηκε για να παρέχει μια αυστηρή γλώσσα με σαφή και ακριβή σύνταξη. Με χρήση της λογικής επιτυγχάνεται η επιβολή αυστηρών κανόνων με σκοπό την αποτελεσματικότερη συλλογή δεδομένων που απαιτούνται για την καλύτερη εξαγωγή αποτελεσμάτων. Παράλληλα αποτελεί το πεδίο που μελετά τις αρχές της συλλογιστικής, προσφέροντας μια κατανοητή τυπική σημασιολογία (formal semantic), διευκολύνοντας έτσι τα προγράμματα της συλλογιστικής να μπορούν να εξάγουν συμπεράσματα. Τέλος η λογική συντελεί στην αποσαφήνιση και τυποποίηση της ανθρώπινης σκέψης. Στόχος είναι μέσω αυτής της διαδικασίας να τεθούν σε εφαρμογή οι κανόνες που παράγονται σε προγράμματα που συλλέγουν τα δεδομένα, και να είναι σε θέση να αναλύσουν δεδομένα και να παράγουν νέα συμπεράσματα. Η βαρύτητα της συγκεκριμένης έννοιας είναι σημαντική για την κατανόηση και τη διαχείριση μεγάλων όγκων δεδομένων, όπως στον Σημασιολογικό Ιστό και στα ανοιχτά δεδομένα. Πιο συγκεκριμένα η διαδικασία αυτή παρέχει στις μηχανές τη δυνατότητα να αξιολογούν κατά την επεξεργασία των δεδομένων τις σχέσεις μεταξύ τους παρέχοντας καλύτερα αποτελέσματα.

2.4.1.2 Πράκτορες

Οι πράκτορες αναφέρονται σαν προγράμματα λογισμικού οπού αναπτυχθήκαν κυρίως με αντικειμενοστρέφεια και η ανάπτυξη τους βασίζεται σε συστατικά μέρη (component-based software development). Το λογισμικό αυτό (πράκτορας) δέχεται ορίσματα από έναν χρήστη και θα αναζητήσει της πληροφορίες από τον Web, θα επικοινωνήσει και με άλλους πράκτορες, θα κάνει τις κατάλληλες επιλογές οπού θα συγκρίνει τις απαιτήσεις του χρήστη και θα επιστρέφει τις απαντήσεις πίσω στον χρήστη. Για την λειτουργία όλου αυτού που περιγράψαμε οι πράκτορες θα πρέπει να χρησιμοποιήσουν όλες τις παραπάνω λειτουργίες .

2.5 Ανοιχτά Δεδομένα και Σημασιολογικός Ιστός

Παρόλα αυτά, πολλά από τα δεδομένα που είναι διαθέσιμα είναι σε μορφές όπως Word, PDF, Excel και αρχεία CSV που προορίζονται κυρίως για ανθρώπινη χρήση καθώς είναι δύσκολα να κατανοηθούν από τις μηχανές, με αποτέλεσμα να οδηγούμαστε συχνά σε άσχετα αποτελέσματα αναζήτησης και αυτό προφανώς περιορίζει τη δυνατότητα εξαγωγής πολύτιμων πληροφοριών. Ο Tim Berners-Lee, ο δημιουργός του Παγκόσμιου Ιστού (World Wide Web), εισήγαγε την έννοια του Σημασιολογικού Ιστού, όπου και αναφέραμε παραπάνω, και αργότερα καθιέρωσε τις αρχές των Συνδεδεμένων Δεδομένων. Έτσι λοιπόν κάνοντας τα δεδομένα αναγνώσιμα από μηχανήματα, οι συστάσεις αυτές βελτιώνουν σημαντικά τον τρόπο με τον οποίο μπορούν να αναζητηθούν, να είναι προσβάσιμα και στην συνέχεια να αξιοποιηθούν τα δεδομένα του παγκόσμιου ιστού. Ο Σημασιολογικός Ιστός δεν είναι απλώς η τοποθέτηση δεδομένων στον ιστό, προσεγγίζει την οργάνωση στη χρήση δεδομένων, αντιμετωπίζοντας τις προκλήσεις που σχετίζονται με την ενοποίηση, τη λειτουργικότητα και την ευρεσιμότητα των δεδομένων χρησιμοποιώντας ετικέτες σε όλο το περιεχόμενο του. Αυτό έχει σαν αποτέλεσμα την επέκταση των δυνατοτήτων καθιστώντας τα δεδομένα όχι μόνο διαθέσιμα αλλά και κατανοητά και αξιοποιήσιμα από τις μηχανές και τα άτομα ώστε να έχουν την δυνατότητα εξερεύνησης στον παγκόσμιο Ιστό.

2.6 Linked Data

Ο ιστός των δεδομένων, όπως και ο ιστός του υπερκειμένου (hypertext), βασίζεται σε έγγραφα που υπάρχουν στον ιστό. Ωστόσο, υπάρχει μια σημαντική διαφορά μεταξύ των δύο. Στον ιστό του υπερκειμένου, οι σύνδεσμοι χρησιμεύουν ως σχεσιακές άγκυρες μέσα σε έγγραφα HTML, συνδέοντας διάφορα κομμάτια πληροφοριών μέσω links που μπορούν να πατηθούν. Από την άλλη πλευρά, στον ιστό των δεδομένων, αυτοί οι σύνδεσμοι συνδέουν αντικείμενα, έννοιες ή οντότητες που περιγράφονται με τη χρήση του (RDF). Αυτές οι συνδέσεις δεν είναι απλώς μεταξύ εγγράφων αλλά μεταξύ σημείων δεδομένων, επιτρέποντας έναν πλουσιότερο και πιο ουσιαστικό ιστό διασυνδεδεμένων δεδομένων.

Ο πυρήνας αυτού του διασυνδεδεμένου ιστού δεδομένων έγκειται στη χρήση των URIs (Uniform Resource Identifiers - URIs). Τα URI λειτουργούν ως αναγνωριστικά για κάθε είδος αντικείμενου ή έννοιας, παρέχοντας έναν τρόπο για την ονομασία και την αναφορά πραγμάτων σε μια τυποποιημένη μορφή που αναγνωρίζεται παγκοσμίως. Αυτό είναι ζωτικής σημασίας, διότι επιτρέπει την απρόσκοπτη ενσωμάτωση και ανάκτηση δεδομένων σε ολόκληρο τον ιστό.



Εικόνα 2.3 Κύκλος Συνδεδεμένων Δεδομένων

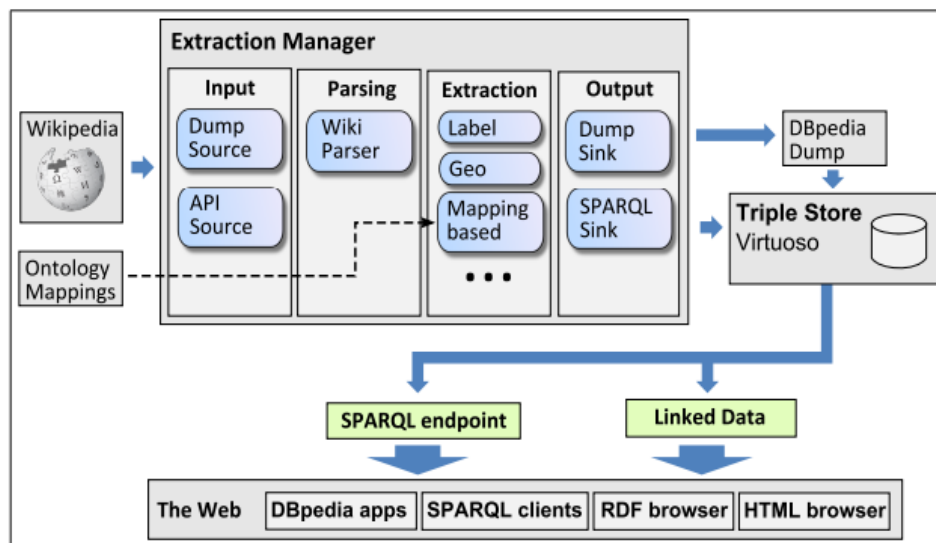
Ο όρος «Διασυνδεδεμένα δεδομένα» αναφέρεται σε ένα σύνολο βέλτιστων πρακτικών για τη δημοσίευση και διασύνδεση δομημένων δεδομένων από διαφορετικές πηγές στον Παγκόσμιο Ιστό. Αυτές οι βέλτιστες πρακτικές είναι γνωστές ως αρχές των συνδεδεμένων δεδομένων που ορίζονται από τον Tim Berners Lee, τον εφευρέτη του Παγκόσμιου Ιστού και του Σημασιολογικού Ιστού, αρχές όπως :

- ✚ **Χρησιμοποιήστε URIs ως ονόματα για τα πράγματα.**
- ✚ **Χρησιμοποιήστε URIs HTTP ώστε οι άνθρωποι να μπορούν να αναζητήσουν αυτά τα ονόματα**
- ✚ **Όταν κάποιος αναζητά ένα URI, παρέχετε χρήσιμες πληροφορίες, χρησιμοποιώντας τα πρότυπα (RDF, SPARQL).**
- ✚ **Συμπεριλάβετε συνδέσμους προς άλλα URIs ώστε να μπορούν να ανακαλύψουν περισσότερα πράγματα.**

2.7 Οφέλη Ανοιχτών Συνδεδεμένων Δεδομένων

Τα συνδεδεμένα ανοικτά δεδομένα είναι μια νέα εξέλιξη όπου όλα τα σύνολα δεδομένων που είναι ελεύθερα διαθέσιμα στο διαδίκτυο συνδέονται με σημασιολογικές τεχνολογίες ώστε να δημοσιεύονται με τρόπο αναγνώσιμο από μηχανήματα και να απλοποιείται η ανάλυσή τους σε μορφή κατανοητή από τον άνθρωπο. Για τον λόγο αυτό μπορούμε να πούμε ότι τα από ανοιχτά συνδεδεμένα δεδομένα έχουμε πολλά οφέλη στα οποία θα μας βοηθούσαν ως προς την επαναχρησιμοποίηση των δεδομένων ακόμα και την αξιοπιστία αυτών ενισχύοντας και άλλο την διαφάνεια. Να αναφέρουμε μερικά πλεονεκτήματα των συνδεδεμένων ανοικτών δεδομένων

- **Αποτελεσματική χρήση των πόρων:** Τα συνδεδεμένα ανοικτά δεδομένα επιτρέπουν σε άτομα και οργανισμούς να επωφεληθούν από δεδομένα που έχουν ήδη συλλεχθεί, επεξεργαστεί και διατεθεί. Αυτό ενθαρρύνει τη συνεργασία και την επαναχρησιμοποίηση των υφιστάμενων πόρων.
- **Βελτίωση της ποιότητας των πληροφοριών:** Τα συνδεδεμένα ανοικτά δεδομένα ενθαρρύνουν την τυποποίηση των μεταδεδομένων και των μορφών δεδομένων, καθιστώντας τα δεδομένα πιο αξιόπιστα και αξιόπιστα.
- **Δημιουργία προστιθέμενης αξίας:** Με την άμεση σύνδεση με άλλα δεδομένα, τα συνδεδεμένα ανοικτά δεδομένα επιτρέπουν στους χρήστες να ανακαλύπτουν, να χρησιμοποιούν και να επαναχρησιμοποιούν πληροφορίες με καινοτόμους τρόπους.
- **Προσδιορισμός των κενών πληροφοριών:** Τα συνδεδεμένα ανοικτά δεδομένα επιτρέπουν την επισήμανση και τη διόρθωση σφαλμάτων στα δεδομένα.
- **Ενίσχυση της διαφάνειας:** Τα συνδεδεμένα ανοικτά δεδομένα είναι προσβάσιμα σε όλους.

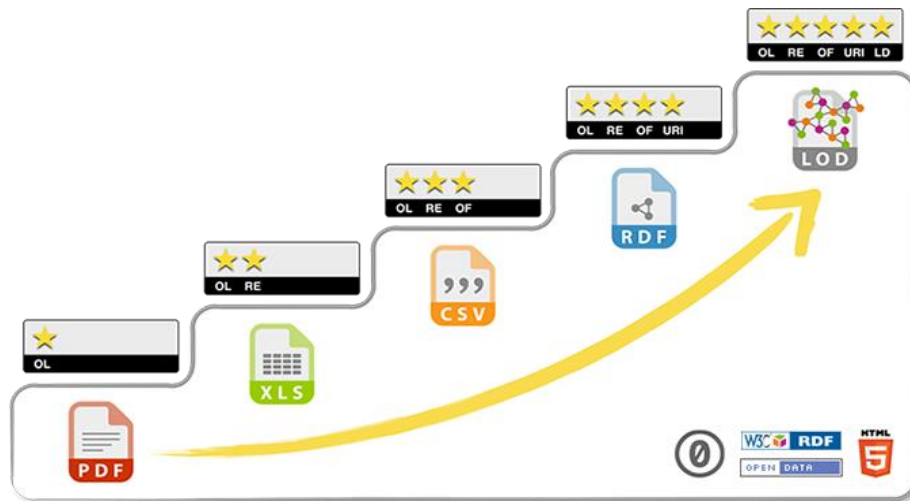


Εικόνα 2.4 Πλαίσιο Εξαγωγής DBpedia

Ένα παράδειγμα **Linked Open Data (LOD)** είναι το **DBpedia**, ένα έργο που εξάγει δομημένα δεδομένα από την Wikipedia και τα συνδέει με άλλες πηγές δεδομένων στον ιστό. Το DBpedia χρησιμοποιεί τα δεδομένα της Wikipedia (όπως ημερομηνίες, τοποθεσίες, πρόσωπα, και γεγονότα) και τα δημοσιεύει ως Linked Open Data, επιτρέποντας τη σύνδεση και την επαναχρησιμοποίηση αυτών των δεδομένων από εφαρμογές και ερευνητές σε όλο τον κόσμο

2.8 5 Star Linked Data

Το σύστημα αξιολόγησης που πρότεινε ο Tim Berners-Lee για τα ανοιχτά συνδεδεμένα δεδομένα (5-star rating for open data) είναι ένας από τους πιο γνωστούς τρόπους για την εκτίμηση της διαφάνειας και της ποιότητας των δεδομένων που διατίθενται δημόσια. Το σύστημα αυτό βασίζεται σε μια κλίμακα αξιολόγησης από 1 έως 5 αστέρια, όπου στο ανώτατο επίπεδο (5 αστέρια) " και κάθε πρόσθετο αστέρι προϋποθέτει ότι τα δεδομένα πληρούν τα κριτήρια του προηγούμενου βήματος ή των προηγούμενων βημάτων. Είναι μια απλή και λογική μέθοδος που βοηθά στην αξιολόγηση της ποιότητας των δεδομένων που δημοσιεύουν κυβερνητικοί φορείς.



Εικόνα 2.5 Διαδρομή προς τα 5 ★ Linked Data

- ★ Διάθεση των δεδομένων στον Ιστό κάτω από μια ανοικτή άδεια με αδιάφορη μορφή
- ★ ★ Διάθεση των δεδομένων ως δομημένα δεδομένα (π.χ., σε μορφή Excel αντί για εικόνα)
- ★ ★ ★ Χρησιμοποίηση των δεδομένων ως μη ιδιόκτητες μορφές (όχι κλειστά πρότυπα) (π.χ., CSV αντί για Excel)
- ★ ★ ★ ★ Επιπλέον Χρήση και προσθήκη URI για ταυτοποίηση, ώστε ο κόσμος να μπορεί να δείχνει τα δεδομένα σας (RDF, SPARQL)
- ★ ★ ★ ★ ★ Σύνδεση δεδομένων με άλλα δεδομένα ώστε να παραχθεί το σημασιολογικό πλαίσιο

Κεφάλαιο 3ο: Τα ανοιχτά Δεδομένα Στην Ελλάδα και στην ΕΕ

3.1 ΕΙΣΑΓΩΓΗ

Η Ελεύθερη και απεριόριστη διάθεση των πληροφοριών και δεδομένων στο εύρη κοινό που επιτρέπει στον καθένα να τις χρησιμοποιεί, να τις τροποποιήσει και να της μοιράζεται, συνήθως παρέχονται από κυβερνήσεις ή δημόσιους φορείς. Γενικά, τα ανοιχτά δεδομένα αποτελούν βασικό μέρος του ψηφιακού μετασχηματισμού στην Ευρώπη και στοχεύει την προώθηση της καινοτομίας, την αύξηση της διαφάνειας και την ενίσχυση της παροχής δημόσιων υπηρεσιών.

Η Ευρωπαϊκή Ένωση (ΕΕ) πρωταγωνιστεί στην πρωτοβουλία για τα ανοιχτά δεδομένα, καθορίζοντας νομοθετικά πλαίσια και χορηγώντας με χρηματοδότηση και πόρους για τη στήριξη της ανάπτυξης ανοικτών δεδομένων σε όλα τα κράτη μέλη της.

Τα ανοιχτά δεδομένα στην Ελλάδα επικεντρώνονται κυρίως σε τομείς όπως η δημόσια διοίκηση, η οικονομική διαφάνεια, τα γεωχωρικά δεδομένα και η δημόσια υγεία. Η δημοσιοποίηση συνόλων δεδομένων έχει επιτρέψει καινοτομίες σε τομείς όπως ο τουρισμός, ο αστικός σχεδιασμός και η διαχείριση καταστροφών. Ωστόσο, εξακολουθούν να υπάρχουν προκλήσεις όπως η ποιότητα των δεδομένων, η πληρότητα και η πολιτισμική αντίσταση στους δημόσιους οργανισμούς. Παρ' όλα αυτά, τόσο η ΕΕ όσο και η Ελλάδα συνεχίζουν να αναπτύσσουν υποδομές και πολιτικές για να ξεπεράσουν αυτά τα εμπόδια, με στόχο την αξιοποίηση των ανοικτών δεδομένων για την οικονομική ανάπτυξη, τη συμμετοχή των πολιτών και τη βελτίωση των δημόσιων υπηρεσιών.



Εικόνα 3.1 data.europa.eu

3.2 ΟΔΗΓΙΑ PSI (PSI Directive (EU))

Τα δεδομένα που είναι ελεύθερα και ευρέως προσβάσιμα (Ανοιχτά δεδομένα) αποτελούν πολύτιμο εργαλείο για την ενίσχυση της οικονομικής και κοινωνικής αξίας. Η οδηγία σχετικά με τα ανοιχτά δεδομένα και την περαιτέρω χρήση πληροφοριών του δημόσιου τομέα διασφαλίζει την κυβερνητική διαφάνεια και τον θεμιτό αγώνα για τους ενδιαφερόμενους που επιθυμούν να επαναχρησιμοποιήσουν δεδομένα του δημόσιου τομέα.

Το 2003, η Ευρωπαϊκή Επιτροπή ενέκρινε ένα νομικό πλαίσιο για την περαιτέρω χρήση των δημόσιων πληροφοριών (PSI), το οποίο αργότερα αναθεωρήθηκε από την Οδηγία 2013/37/ΕΕ. Μέχρι το 2030, η συνολική άμεση οικονομική αξία του PSI προβλέπεται να ανέλθει στα 194 δισεκατομμύρια ευρώ σύμφωνα με το European Commission. Η οδηγία του 2013 επικεντρωνόταν περισσότερο στις οικονομικές πτυχές της επαναχρησιμοποίησης των πληροφοριών, μέχρι ως που μια αναθεώρηση του 2019 που είχε στόχο την οικονομική ανάπτυξη και την καινοτομία νέων προϊόντων και νέων υπηρεσιών, να επιτευχθούν άμεσες καινοτόμες λύσεις στις κοινωνικές προκλήσεις, όπως για παράδειγμα στην υγεία

και στις κοινωνικές μεταφορές την βελτίωση της διαφάνειας, την ενίσχυση στην πολιτική συμμετοχή των πολιτών καθώς και την ενίσχυση της Τεχνικής Νοημοσύνης (AI).

Έτσι λοιπόν η οδηγία για τα ανοικτά δεδομένα και την ελεύθερη χρήση πληροφοριών του δημόσιου τομέα, γνωστή και ως οδηγία για τα ανοικτά δεδομένα, τέθηκε σε ισχύ στις 16 Ιουλίου 2019, αντικαθιστώντας την οδηγία του 2003 για τις δημόσιες πληροφορίες. Σύμφωνα με την οδηγία (ΕΕ) 2019/1024 , τα δεδομένα του δημόσιου τομέα που είναι διαθέσιμα βάσει των εθνικών νόμων περί ελευθερίας της πληροφόρησης μπορούν να επαναχρησιμοποιηθούν, με ελεύθερη χρέωση ή χρέωση που δεν υπερβαίνει το οριακό κόστος.

Στα σύνολα που τα δεδομένα έχουν υψηλή αξία τα σύνολα τους θα πρέπει να διατίθενται δωρεάν αφού η επαναχρησιμοποίηση προσφέρουν σημαντικά οφέλη για την κοινωνία και την οικονομία, όπως έχουμε αναλύσει και στις προηγούμενες ενότητες ,αυτά τα σύνολα δεδομένων πρέπει να διατίθενται δωρεάν, σε μορφότυπους αναγνώσιμους από μηχανήματα και μέσω API και μαζικής λήψης .Οι κατηγορίες των συνόλων δεδομένων υψηλής αξίας, όπως αναφέρονται στο άρθρο 13 παράγραφος 1 της οδηγίας, είναι οι γεωχωρικές , γεωσκοπήσεις και περιβάλλον , μετεωρολογικές , στατιστικές ,εταιρείες και ιδιοκτησίες εταιρειών ,κινητικότητα

3.3 Ευρωπαϊκή Στρατηγική

Η οδηγία για τα ανοικτά δεδομένα στην Ευρώπη προβλέπει κοινούς κανόνες για μια ευρωπαϊκή αγορά δεδομένων που βρίσκονται στην κατοχή του Δημοσίου. Η Ευρωπαϊκή Επιτροπή έχει ως στόχο την ενίσχυση της ως προς στα ψηφιακά δεδομένα

Η στρατηγική της Ευρωπαϊκής Ένωσης, για τα δεδομένα της αποσκοπεί στη δημιουργία μιας ενιαίας αγοράς δεδομένων, ενισχύοντας την ανταγωνιστικότητα σε παγκόσμιο επίπεδο και διασφαλίζοντας την αξία των δεδομένων κυρίως δίνει έμφαση στην υπεύθυνη και ασφαλή χρήση των δεδομένων για την προώθηση της οικονομικής ανάπτυξης, της καινοτομίας και της δημιουργίας θέσεων εργασίας. Η στρατηγική τονίζει την αξία των δεδομένων για τη βελτίωση της υγειονομικής περιθάλψης, την ανάπτυξη ασφαλέστερων συστημάτων μεταφορών, τη δημιουργία νέων προϊόντων και υπηρεσιών και την προώθηση της βιωσιμότητας. Σύμφωνα με την επιτροπή της Ευρώπης με την δημιουργία κοινής αγοράς για τα δεδομένα αυτά θα μπορούν ρέουν ελεύθερα εντός της ΕΕ και σε όλους τους τομείς , με προϋποθέσει να τηρούνται όλοι οι κανόνες οι ευρωπαϊκοί ιδίως της προστασίας της ιδιωτικής ζωής και των δεδομένων καθώς και στην ισότητα των όλων για την πρόσβαση στα δεδομένα προς όφελος των επιχειρήσεων , ερευνητών και δημόσιων διοικήσεων

Σε ισχύει από τον Ιανουάριο του 2024 τέθηκε η ευρωπαϊκή πράξη για τα δεδομένα στοχεύοντας να επιτρέψει την μεγαλύτερη χρήση των δεδομένων στην ΕΕ. Οι νέοι κανόνες αυτοί προβλέπονται να δημιουργήσουν αύξηση του ΑΕΠ ύψος 270 δισεκατομμύρια ευρώ για τα κράτη της ΕΕ έως το 2028

Επίσης να προσθέσουμε ότι η Ευρωπαϊκή Επιτροπή έχει δημιουργήσει την **Ευρωπαϊκή Πύλη Δεδομένων (European Data Portal - EDP)**, μέσω της οποίας συγκεντρώνονται και διατίθενται τα μεταδεδομένα από τα κεντρικά αποθετήρια ανοιχτών δημόσιων δεδομένων των κρατών μελών καθώς και κάθε χώρα της ευρωπαϊκής ένωσης χειιάζετε να λειτουργεί μια πύλη ανοιχτών δεδομένων.

Τα προβλεπόμενα αριθμητικά στοιχεία για το 2025 είναι :

- **Αύξηση 530% όγκου δεδομένων**
- **829 ΑΕΠ Δισ. Ευρώ**
- **Αύξηση των Επαγγελματιών σε 10,9 Εκατομμύρια από 5,7 το 2018**
- **Αύξηση βασικών ψηφιακών δεξιοτήτων του πληθυσμού της ΕΕ κατά 65 %**



Εικόνα 3.2 Προβλεπόμενα αριθμητικά στοιχεία για το 2025

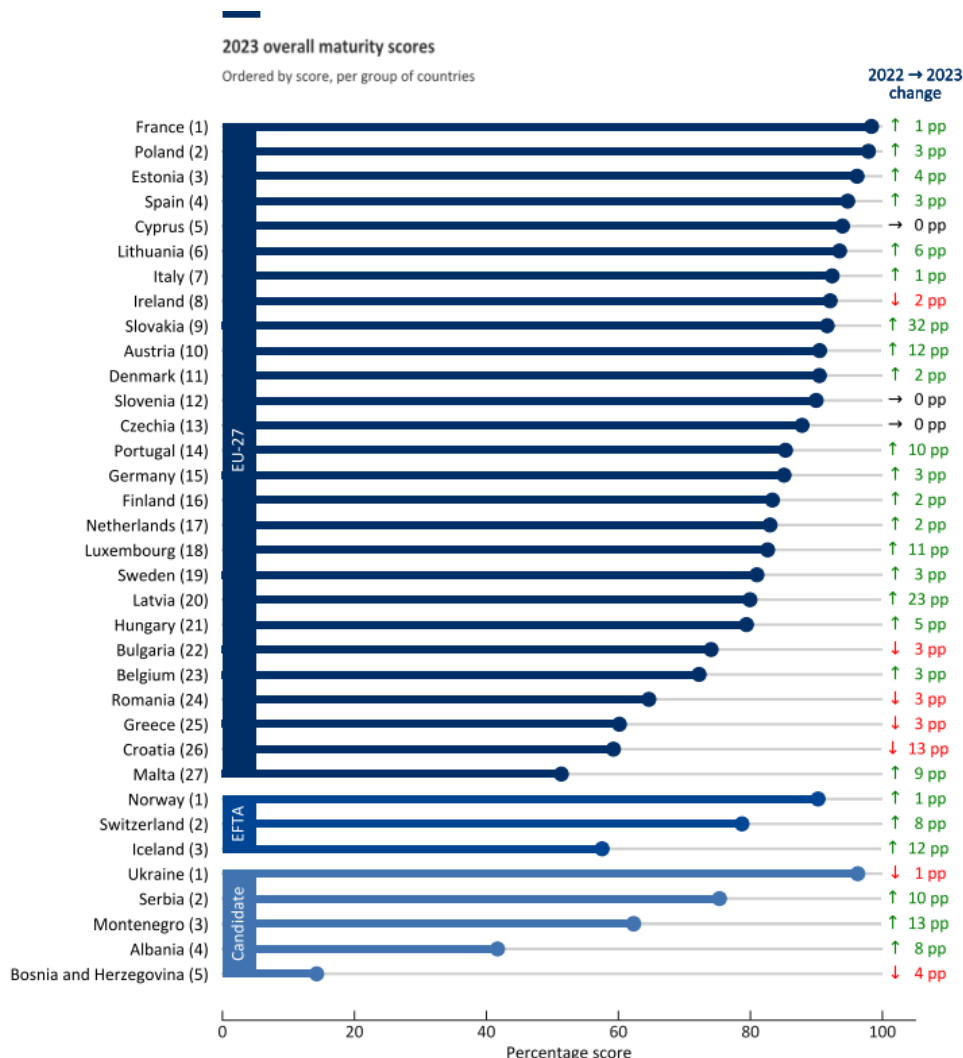
3.4 Ανοιχτά Δεδομένα στην Ελλάδα

Τα ανοιχτά δεδομένα αποτελούν έναν από τους βασικούς πυλώνες για την ενίσχυση της διαφάνειας και την υποστήριξη της καινοτομίας σε κάθε χώρα. Στην Ελλάδα, τα τελευταία χρόνια έχουν γίνει σημαντικές προσπάθειες από την κυβέρνηση και διάφορους οργανισμούς για τη βελτίωση της πρόσβασης σε δεδομένα που συλλέγονται από δημόσιους φορείς. Κυρίως μέσω της Εθνικής Πύλης Ανοιχτών Δεδομένων (data.gov.gr), οι πολίτες, οι επιχειρήσεις και οι ερευνητές έχουν τη δυνατότητα να αξιοποιήσουν δεδομένα σε διάφορους τομείς, όπως η υγεία, η εκπαίδευση, το περιβάλλον και τα οικονομικά.

Σύμφωνα με την Αναφορά του **Open Data Maturity Report 2023**, που αξιολογεί την ωριμότητα των ανοιχτών δεδομένων σε όλη την Ευρώπη, η Ελλάδα παρουσιάζει πρόοδο, αλλά βρίσκεται πίσω από πολλές ευρωπαϊκές χώρες. Η αξιολόγηση των χωρών βασίζεται σε τέσσερις δείκτες όπως :

- 1) **Open Data Policy (Πολιτική Ανοιχτών Δεδομένων):** αφορά το νομικό και στρατηγικό πλαίσιο που
- 2) **Open Data Impact (Επίδραση Ανοιχτών Δεδομένων):** Η επίδραση που έχουν τα ανοιχτά δεδομένα στην οικονομία, την κοινωνία και τη δημόσια διοίκηση.
- 3) **Open Data Portal (Πύλη Ανοιχτών Δεδομένων):** Ο δείκτης αυτός αξιολογεί την ποιότητα και την ευχρηστία των πλατφορμών που παρέχουν πρόσβαση σε ανοιχτά δεδομένα
- 4) **Open Data Quality (Ποιότητα Ανοιχτών Δεδομένων):** Αφορά την ποιότητα των δεδομένων, τη διαθεσιμότητα ενημερωμένων και ολοκληρωμένων συνόλων δεδομένων και τη δυνατότητα δια λειτουργικότητας

Συγκριτικά με τις υπόλοιπες χώρες της Ευρωπαϊκής Ένωσης, σύμφωνα πάντα με το Open Data Maturity Report 2023 η Ελλάδα βρίσκεται χαμηλά στον πίνακα κατανέμοντας στην 25 θέση, με μέσο όρο ευρωπαϊκής ωριμότητας ανοιχτών δεδομένων με περίπου στο 60 % κάτω από τον μέσο όρο της Ευρώπης Ένωσης είναι στο 83% από 79% που ήταν το 2022, όπως θα δούμε και παρακάτω η θέση της Ελλάδας βρίσκεται πάνω μόνο την Μάλτα και την Κροατία, παρουσιάζοντας μια μικρή πτώση σε σχέση με τις μετρήσεις του 2022



Πίνακας 3.1 Συνολικές βαθμολογίες ωριμότητας για όλες τις συμμετέχουσες χώρες το 2023, με κατάταξη από την υψηλότερη βαθμολογία προς τη χαμηλότερη βαθμολογία(
https://data.europa.eu/sites/default/files/odm2023_report.pdf)

Στην χώρα μας μπορούμε να πούμε έχουμε περιθώρια βελτιώσεις καθώς η χώρα ακόμα διανύει τη διαδικασία βελτίωσης των δομών, των μηχανισμών και των πρακτικών που θα επιτρέψουν τη μέγιστη αξιοποίηση των δεδομένων αυτών για δημόσιο όφελος και αυτό φαίνεται από τον πίνακα της στρατηγικής πολιτικής για τα ανοιχτά δεδομένα

3.4.1 Data.gov.gr

Σύμφωνα με τους κανονισμούς της Ευρωπαϊκής ένωσης πρέπει όλες οι χώρες που ανήκουν σε αυτοί να συμμορφωθούν σύμφωνα με τους κανονισμούς, έτσι λοιπόν και η Ελλάδα έχει αναπτύξει το data.gov.gr για την παροχή των ανοιχτών δεδομένων του ελληνικού δημοσίου. Πρόκειται για μια αναπτυγμένη πλατφόρμα που παρέχει δεδομένα από διάφορους τομείς της δημόσιας διοίκησης, όπως μετακινήσεις, μετεωρολογικά, δημογραφικά στοιχεία και εγκληματικότητα ανοιχτά, οργανωμένα σε 74 σειρές δεδομένων και 10 θεματικές κατηγορίες δημιουργώντας για πρώτη φορά μια συγκέντρωση δεδομένων αξιοποιώντας της σύγχρονες τεχνολογίες. Τα δεδομένα από το data.gov.gr παρέχονται μέσω API στο κοινό που επιθυμούν να κάνουν λήψη τα δεδομένα αυτά χρησιμοποιώντας απλά ένα Token που τους αποστέλλεται πχ. μέσω email για την λήψη των δεδομένων.

Η ενέργεια αυτή ενσωματώνετε σε συνέχεια της της Οδηγίας (ΕΕ) 2019/1024 (βλ. Ενότητα 7.2) για τα ανοιχτά δεδομένα, μέσω του Νόμου 4727/2020 για την Ψηφιακή Διακυβέρνηση και τις Ηλεκτρονικές Επικοινωνίες.

Να σημειωθεί ότι η Ελλάδα αποτελεί την πρώτη χώρα που η συγκεκριμένη οδηγία ενσωματώνεται στο εθνικό δίκαιο της χώρας εκσυγχρονίζοντας το νομοθετικό πλαίσιο 4305/2014

Στην χώρα μας γίνεται μια αξιολόγηση προσπάθεια ενσωμάτωσης των ανοιχτών δεδομένων πέρα από το data.gov.gr οπου είναι και η κυριά πύλη του Ελληνικού Δημοσίου οπου μπορούμε να βρούμε διαφορά δεδομένα από διάφορες κατηγορίες θεματικές ενότητες συνόλων δεδομένων όπως και άλλωστε αναφέραμε στην προηγούμενη παράγραφο οι οποίες είναι οι παρακάτω:

-  **Δημοσία Τάξη και Δικαιοσύνη**
-  **Εκπαίδευση**
-  **Κοινωνία**
-  **Κράτος και Δημοκρατία**
-  **Μετακινήσεις**
-  **Οικονομία και Εμπόριο**
-  **Περιβάλλον και Ενέργεια**
-  **Τεχνολογία**
-  **Τηλεπικοινωνίες**
-  **Υγεία**

Έτσι λοιπόν κάτω από κάθε κατηγορία περιέχονται τα σύνολα δεδομένων μερικά με τα πιο Δημοφιλέστερα σύνολα δεδομένων με βάση το Data.gov.gr είναι όπως

- a) **Αριθμός Κρατουμένων Σωφρονιστικών Καταστημάτων**
- b) **Μητρώο Πολιτών**
- c) **Αριθμός Μαθητών ανά Σχολείο**
- d) **Μετεωρολογικά Δεδομένα Μετεωρολογικών Σταθμών Κρήτης**
- e) **Επιβατικό κοινό στον ΟΑΣΑ**
- f) **Κυκλοφορία διαδικτύου στην Ελλάδα**
- g) **Παράλληλες Εξαγωγές Φαρμάκων**
- h) **Δαπάνες Ε&Α ανά πηγή χρηματοδότησης**

και επιπλέον αναφέρονται επιγραμματικά το πως είναι χωρισμένα τα δεδομένα τα πεδία τους οι στήλες και οτιδήποτε άλλο.

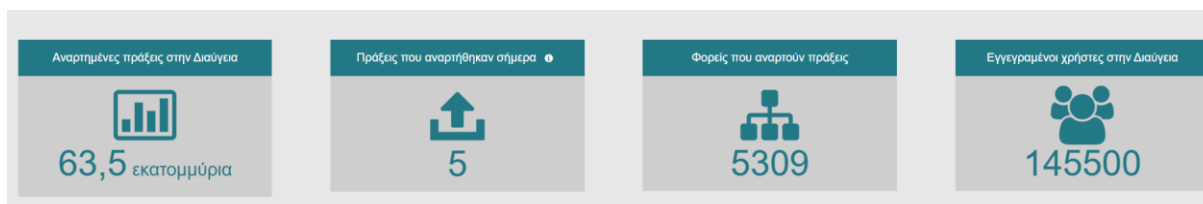
Στην Ελλάδα υπάρχουν αρκετές πύλες δεδομένων, οι οποίες προσφέρουν ανοιχτά δεδομένα για διάφορους τομείς. Εκτός από το **govData**, που ήδη αναφέρθηκε, οι δύο από τις πιο σημαντικές και αναπτυγμένες πλατφόρμες στην χώρα μας είναι η **Διαύγεια** και η **Geodata** (γίνετε παρακάτω αναφορά για τις δυο αυτές πύλες δεδομένων.)

Εκτός από αυτές τις πλατφόρμες, υπάρχουν μικρότερες πύλες που εστιάζουν σε τοπικές περιοχές, όπως πχ. Open Data Κρήτης(<https://data.apdkritis.gov.gr/el>), Open Data Αττικής (<http://www.opendata.attica.gov.gr>), Open Data Θεσσαλονίκης (<https://opendata.thessaloniki.gr/>) κ.α

3.4.2 Διαύγεια

Τον Οκτώβριο του 2010 άρχισε η λειτουργία της πύλης Διαύγεια οπού στοχεύει την δημοσιά κυβερνητική πολιτική και την διοικητική δραστηριότητα διασφαλίζοντάς την διαφάνεια και την εμπέδωση της υπευθυνότητας και της λογοδοσίας σύμφωνα με τον νομό 3861/2010 οπού όλες οι αναρτήσεις αποφάσεων κυβερνητικών και διοικητικών οργάνων αναρτώνται σε μια πύλη αναρτώνται οι αποφάσεις των πράξεων.

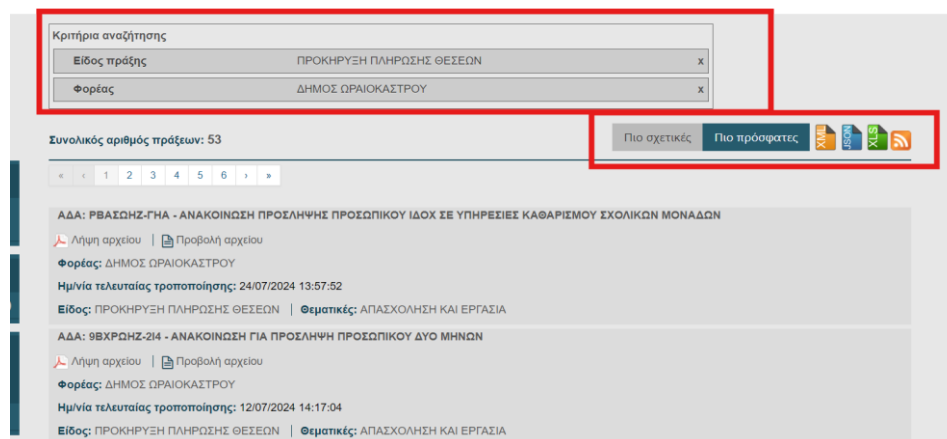
Μέχρι σήμερα στην πύλη αυτή περιλαμβάνονται 5309+ φορείς δημοσίου τομέα, συμπεριλαμβανομένων των Οργανισμών Τοπικής Αυτοδιοίκησης πρώτου (Α) και δεύτερου (Β) βαθμού και έχουν αναρτηθεί παραπάνω από 63,5 εκατομμύρια πράξεις (πράξεις εννοείται οποιοδήποτε επίσημο έγγραφο ή απόφαση που εκδίδεται από φορείς του δημόσιου τομέα) και 145500 εγγεγραμμένοι χρήστες.



Εικόνα 3.3 Πλήθος Δεδομένων Διαύγεια

Κάθε πράξη αποκτά ισχύει μόνο μέσω αναστήσεις στο Σύστημα της Διαυγείας , υπογράφεται ψηφιακά από το σύστημα και κάθε πράξη παίρνει έναν μοναδικό Αριθμό Διαδικτυακής Ανάρτησης (ΑΔΑ) -το ID (ταυτότητα) κάθε πράξης ,το έγγραφο που αναρτιέται στο σύστημα υπερισχύει όλων τον άλλων εκδόσεων της πράξης εκτός από όσα δημοσιεύονται στην Εφημερίδας της Κυβέρνησης . Επίσης ολοι οι πολίτες μπορούν να έχουν πρόσβαση στα έγγραφα ελεύθερα και όσο για της υπηρεσίες μπορούν να διεκπεραιώνουν υποθέσεις χωρίς να διακινούν αναρτημένα έγγραφα, όλα αυτά ενισχύονται θεσμικά με την ψήφιση του Ν. 4210/2023.

Ένα σχετικά Γρήγορο παράδειγμα ότι ένας πολίτης μπορεί να ψάξει για να δει τις πρόσφατες αναρτήσεις πρόσληψης υπαλλήλων από τον δήμο της περιοχής του πχ.(Δήμος Ωραιοκάστρου), έτσι μπορεί να ελέγξει πόσες προσλήψεις έγιναν, σε ποιες θέσεις, και αν τηρήθηκαν οι διαδικασίες διαφάνειας και αξιοκρατίας.



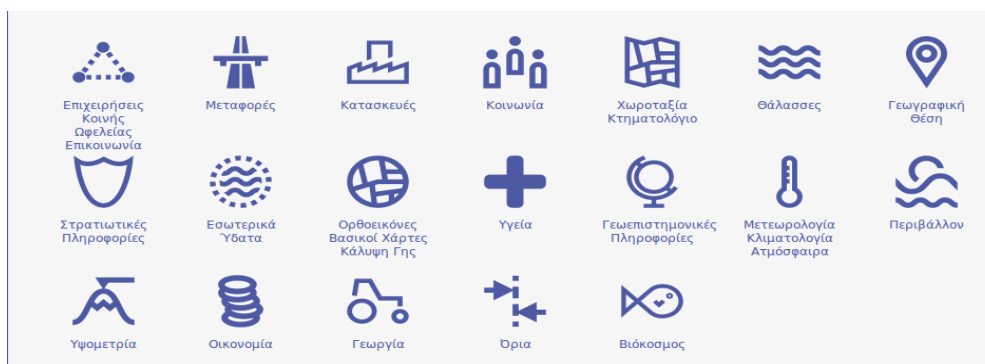
Εικόνα 3.4 Εικόνα παραδείγματος με Προσλήψεις Διαύγεια

Όπως βλέπουμε και στην παραπάνω εικόνα έχουμε αναζητήσει πράξεις που αφορούν Προκηρύξεις Πλήρωσης Θέσεων με φορέα τον Δήμο Ωραιοκάστρου. Έχουμε τη δυνατότητα να δούμε κάθε πράξη σε μορφή **PDF**, που περιλαμβάνει το επίσημο έγγραφο της έγκρισης που έχει δοθεί. Αυτό είναι χρήσιμο για την άμεση ανάγνωση και επιβεβαίωση της απόφασης όπως δημοσιεύτηκε και επιπλέον μέσω των αρχείων XML,JSON η XLS που μας δίνει η πύλη αυτή μπορούν πολύ ευκολά ερευνητές, δημοσιογράφοι ή προγραμματιστές να χρησιμοποιήσουν αυτά τα δεδομένα για να αναπτύξουν εφαρμογές ,να κάνουν αναλύσεις σε μαζικά δεδομένα, ή να εξάγουν συγκεκριμένες πληροφορίες εύκολα και γρήγορα χωρίς να χρειάζεται να διαβάσουν κάθε πράξη μία προς μία σε μορφή PDF.

3.4.3 GEODATA

Το Geodata.gov.gr είναι μια άλλη ανοικτή πλατφόρμα που μοιράζεται γεωχωρικά δεδομένα και υπηρεσίες στην Ελλάδα, λειτουργώντας τόσο ως εθνικός κατάλογος ανοικτών δεδομένων όσο και ως συμβατή υποδομή γεωχωρικών πληροφοριών σύμφωνα με τα πρότυπα INSPIRE.

Δημιουργήθηκε το 2010 και ήταν ένας από τις πρώτες πύλες ανοικτών δεδομένων στον κόσμο, προωθώντας την ανοικτή διακυβέρνηση, όπου οι χρήστες έχουν την δυνατότητα να δημοσιεύσουν , να διερευνούν και να επαναχρησιμοποιούν τα γεωχωρικά δεδομένα με ευκολία. Το GeoData βασίζεται αποκλειστικά σε ελεύθερο λογισμικό και λογισμικό ανοικτού κώδικα, αξιοποιώντας το έργο PublicaMundi και της παγκόσμιας κοινότητας ανοικτού κώδικα. Παρέχει ευκολότερη διανομή και δημοσίευση γεωχωρικών δεδομένων με αυτή την πλατφόρμα, οι χρήστες μπορούν να ανεβάζουν σύνολα δεδομένων σε λιγότερο από ένα λεπτό έτοιμα για επαναχρησιμοποίηση



Εικόνα 3.5 Γεωχωρικά Θέματα από την Geodata.gov.gr

Επίσης, έχει διαθέσιμες Ανοιχτές Διαδικτυακές υπηρεσίες και API δίνοντας σε όλους να τα εφαρμόσουν στις εφαρμογές τους, παρέχοντας έτσι εφαρμογές όπως

1. **Υπηρεσίες OGC (Open Geospatial Consortium) περιλαμβάνει πρότυπα σχετικά με διεπαφές.**
2. **API Δεδομένων**
3. **Χαρτογραφικό API**
4. **Διασυνδεδεμένα δεδομένα**

Οπού τα γεωγραφικά δεδομένα απλοποιούνται ανεξαρτήτως αν υπάρχουν γνώσεις για τα γεωχωρικά η όχι η αν θα χρησιμοποιηθεί από κάποιο mobile App η οτιδήποτε άλλο υπάρχουν λύσεις για τα πάντα , οι εφαρμογές είναι επεκτάσιμες και γρήγορες επίσης καθώς έχουν σχεδιάσει να εξυπηρετήσουν συνθέτους ς φόρτους εργασίας μεγάλων δεδομένων (Big Data) και το API είναι για όλους αρκεί να επιλεγεί ένα από τα πολλά είτε για την προσθήκη ενός χάρτη είτε και για πολλά άλλα πράγματα.

Η πύλη αυτή είναι μια εξαιρετικά χρήσιμη και καλοσχεδιασμένη υπηρεσία, που προσφέρει πρόσβαση σε γεωχωρικά δεδομένα για την Ελλάδα, αποτελεί ένα ισχυρό εργαλείο για επαγγελματίες, ερευνητές και πολίτες που ενδιαφέρονται για την ανάλυση και τη χρήση χωρικών δεδομένων. Ωστόσο, παρά τις δυνατότητές της, δεν έχει αξιοποιηθεί στο μέγιστο βαθμό, καθώς πολλά από τα διαθέσιμα δεδομένα είναι παλαιά ή δεν ενημερώνονται συχνά και η ανάρτηση νέων δεδομένων δεν πραγματοποιείτε συχνά

3.5 Η Δράση του Open Knowledge Foundation

Από το 2004 οπού το Open Knowledge Foundation (Ίδρυμα Ανοικτής Γνώσης) πρωτοεμφανίστηκε αρχίσει αμέσως να ασχολείται με την διασταύρωση των Ψηφιακών Εργαλείων και διαμορφώνει ένα ευρύ δίκτυο κοινοτήτων για την εξυπηρέτηση του δημόσιου συμφέροντος μέσω της ανοικτής γνώσης «εξοπλίζοντας» τους ανθρώπους με γνώσεις και δεξιότητες δημιουργώντας δίκτυα που κατανοούν τη μετασχηματιστική δύναμη των Ανοικτών δεδομένων. Η ιδέα είναι να διαμορφωθούν μελλοντικές υποδομές και οργανισμούς με θεμελιώδη αρχή την ανοικτότητα, στοχεύοντας σε έναν κόσμο δίκαιο, βιώσιμο και χωρίς αποκλεισμούς για όλους για την βοήθεια της αντιμετώπισης σοβαρών προκλήσεων όπως η φτώχεια , κλιματική αλλαγή και αλλά.



Εικόνα 3.6 Λογότυπο του Open Knowledge Foundation

Όπως και σε πολλές χώρες η δράση του Open Knowledge Greece δραστηριοποιείτε και στην Ελλάδα από το 2013 ως επίσημο παράρτημα του Open Knowledge International, παράγοντας διαφορά ερευνητικά έργα και εφαρμογές οπού θα αναπτύσσονται λύσεις που επιδρούν θετικά στην κοινωνία επικεντρώνοντας στην προώθηση της διαφάνειας και της ανοικτής πρόσβασης στα δεδομένα βελτιώνοντας σημαντικά την δημοσία πληροφορία που είναι χρήσιμη για την κοινωνία.

Κεφάλαιο 4ο: Πύλες Ανοιχτών Δεδομένων

4.1 ΕΙΣΑΓΩΓΗ

Τα ανοικτά δεδομένα έχουν αναδειχθεί ως ένας βασικός πόρος στη σημερινή κοινωνία της πληροφορίας, προσφέροντας απεριόριστη πρόσβαση σε δεδομένα που μπορούν να χρησιμοποιηθούν, να τροποποιηθούν και να διαμοιραστούν ελεύθερα από οποιονδήποτε για οποιονδήποτε σκοπό. Η ευρεία διαθεσιμότητα των ανοικτών δεδομένων προάγει τη διαφάνεια, ενθαρρύνει την καινοτομία και υποστηρίζει τη λήψη τεκμηριωμένων αποφάσεων σε διάφορους τομείς, όπως η κυβέρνηση, η ακαδημαϊκή κοινότητα και η βιομηχανία. Για να αξιοποιηθεί πλήρως το δυναμικό των ανοικτών δεδομένων, είναι σημαντικό να διασφαλιστεί ότι τα δεδομένα δεν είναι μόνο προσβάσιμα αλλά και καλά οργανωμένα, έγκαιρα και αναγνώσιμα από μηχανήματα.

Κεντρικό ρόλο στην αποτελεσματική διάδοση και αξιοποίηση των ανοικτών δεδομένων διαδραματίζουν οι ειδικές πλατφόρμες και πύλες που χρησιμεύουν ως αποθετήρια και σημεία πρόσβασης για ποικίλα σύνολα δεδομένων. Αυτές οι πλατφόρμες πυλών ανοικτών δεδομένων παρέχουν δομημένα περιβάλλοντα όπου τα δεδομένα μπορούν να δημοσιεύονται, να διαχειρίζονται και να προσπελαύνονται αποτελεσματικά.



Εικόνα 4.1 CKAN – DKAN

Διαδραματίζουν κρίσιμο ρόλο στην ενίσχυση της δυνατότητας ανακάλυψης δεδομένων, στη διευκόλυνση της διαλειτουργικότητας και στη δυνατότητα των χρηστών να αντλούν ουσιαστικές γνώσεις μέσω διαφόρων εργαλείων και διεπαφών και APIs. Η επιλογή της κατάλληλης πλατφόρμας - είτε πρόκειται για πλατφόρμα ανοικτού κώδικα όπως το CKAN και το DKAN είτε για εμπορικές υπηρεσίες- επηρεάζει σημαντικά την ποιότητα και την επαναχρησιμοποίηση των δεδομένων.

Οι πλατφόρμες ανοικτού κώδικα, όπως το CKAN, όπου θα αναφερθούμε αρκετά σε αυτό το κεφάλαιο, αλλά και το DKAN, προσφέρουν υψηλή προσαρμογή και ευελιξία μέσω των διαθέσιμων APIs, καθιστώντας τις δημοφιλείς επιλογές για οργανισμούς που διαθέτουν την τεχνική εμπειρία για τη διαχείρισή τους. Το CKAN, για παράδειγμα, χρησιμοποιείται ευρέως από κυβερνήσεις και οργανισμούς σε όλο τον κόσμο λόγω του ισχυρού συνόλου χαρακτηριστικών του και της ισχυρής υποστήριξης της κοινότητας. Το DKAN, μια παραλλαγή του CKAN που βασίζεται στο Drupal, παρέχει πρόσθετες δυνατότητες διαχείρισης περιεχομένου, καθιστώντας το μια ευέλικτη επιλογή για όσους είναι εξοικειωμένοι με το οικοσύστημα του Drupal.

4.1.1 API

Ο όρος API (Application Programming Interface) είναι μια διεπαφή προγραμματιστικών διαδικασιών που παρέχει ένα λειτουργικό σύστημα προκειμένου να γίνονται αιτήσεις (Requests) από αλλά προγράμματα και την απάντηση τους σε αυτά. Διευκολύνει σημαντικά την ανταλλαγή δεδομένων. Μέσω των APIs πλατφόρμες όπως το CKAN και όχι μόνο μπορούν να προσφέρουν με ευέλικτο τρόπο τις λειτουργίες ανταλλαγής δεδομένων ενισχύοντας σημαντικά την διαλειτουργικότητα και της δυνατότητες των εφαρμογών όπως θα δούμε και παρακάτω.

4.1.2 DRUPAL

Το Drupal είναι ένα λογισμικό διαχείρισης περιεχομένου (Content Management System - CMS) ανοικτού κώδικα, γραμμένο σε γλώσσα προγραμματισμού PHP. Το Drupal το καθιστά ξεχωριστό το γεγονός ότι είναι εύκολη η δημιουργία του περιεχομένου με αξιόπιστη απόδοση και υψηλή ασφάλεια παρέχοντας τας απολύτως δωρεάν σύμφωνα με τους όρους της γενικής δημοσίας άδεια GNU(GPL). Έχει ευελιξία παρέχοντάς ένα εύρη φάσμα από modules επιτρέποντας την δημιουργία προσαρμοσμένων λύσεων από διάφορες ανάγκες που μπορεί να προκύψουν.



Εικόνα 4.2 Λογότυπο Drupal

4.2 CKAN

Το CKAN είναι μια πλατφόρμα ανοικτού κώδικα (Open) που έχει σχεδιαστεί για τη δημιουργία συλλογής ανοικτών δεδομένων, παρόμοια με τον τρόπο διαχείρισης περιεχομένου του WordPress, αλλά με έμφαση στα δεδομένα αντί για τα blogs. Επιτρέπει σε οργανισμούς όπως κυβερνήσεις και ερευνητικά ιδρύματα να διαχειρίζονται και να δημοσιεύουν σύνολα δεδομένων. Κάθε σύνολο δεδομένων περιλαμβάνει μεταδεδομένα, όπως ο τίτλος, ο εκδότης, πληροφορίες License (άδειας), καθώς και τα datasets τα οποία μπορεί να είναι σε μορφές όπως CSV, Excel, XML ή PDF.

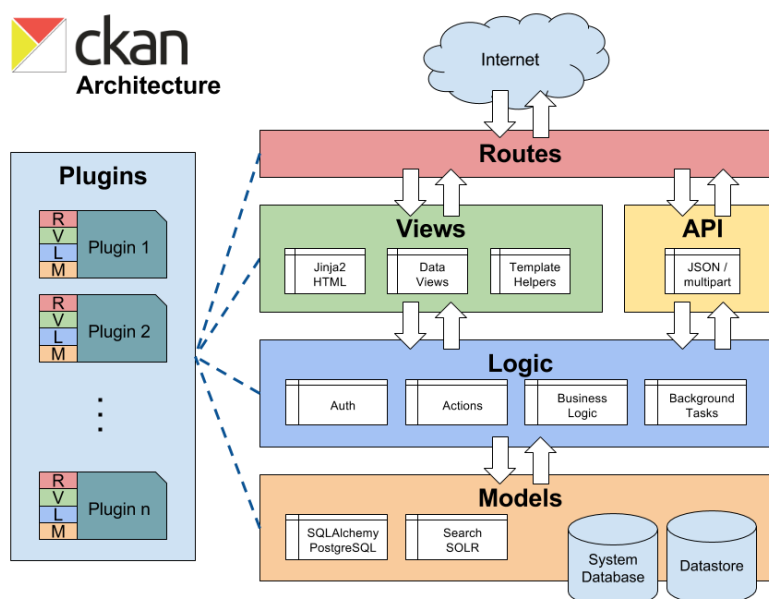
Οι χρήστες μπορούν να περιηγηθούν, να αναζητήσουν και να κάνουν προεπισκόπηση των συνόλων δεδομένων μέσω της φιλικής προς το χρήστη διεπαφή του CKAN, η οποία υποστηρίζει χάρτες, γραφήματα και πίνακες. Ενώ ο καθένας μπορεί να αναζητήσει δεδομένα, μόνο εξουσιοδοτημένοι (administrator) χρήστες μπορούν να δημοσιεύσουν ή να επεξεργαστούν σύνολα δεδομένων, τα οποία συνήθως διαχειρίζονται από συγκεκριμένους οργανισμούς. Αυτοί οι οργανισμοί ελέγχουν τις διαδικασίες δημοσίευσης των δεδομένων τους και μπορούν να αναθέτουν ρόλους στους χρήστες, επιτρέποντας τη συνεργασία και διατηρώντας παράλληλα την ακεραιότητα των δεδομένων.

Το CKAN επιτρέπει επίσης όπου αυτό είναι απαραίτητο την ευελιξία, όπως η δημιουργία δημόσιων συνόλων δεδομένων που μπορούν να επεξεργαστούν από οποιονδήποτε εγγεγραμμένο χρήστη, προωθώντας ένα συνεργατικό περιβάλλον που μοιάζει με ένα datahub.

Συνολικά, το CKAN είναι ένα ισχυρό εργαλείο για κάθε οργανισμό που επιθυμεί να καταστήσει τα δεδομένα του πιο προσβάσιμα και αξιοποιήσιμα, παρέχοντας παράλληλα τους απαραίτητους ελέγχους για τη διαχείριση της δημοσίευσης δεδομένων και των δικαιωμάτων.

4.2.1 Αρχιτεκτονική του CKAN

Ένα από τα πιο ενδιαφέροντα τεχνικά κομμάτια του CKAN είναι το πώς αυτό λειτουργεί δηλαδή ποια είναι η αρχιτεκτονική του και πώς η πύλη αυτή είναι σχεδιασμένη. Ας παρουσιάσουμε για την καλύτερη κατανόηση της αρχιτεκτονική αυτή με ένα παράδειγμα, μπορούμε να πούμε ότι ένας χρήστης επισκέπτεται στην πλατφόρμα και επιθυμεί την ανάκτηση των δεδομένων X , με την χρήση του URL στον browser οι **Διαδρομές (Routes)** ορίζουν την σύνδεση με τις **Προβολές (Views)** όπου ,δέχεται αυτό το αίτημα από τον χρήστη, όπου έχουν την ευθύνη για την επιστροφή της απάντησης αποδίδοντας πρότυπα jinja2 χρησιμοποιώντας και βοηθητικές συναρτήσεις Templates helper που χρησιμοποιούνται για κώδικα που επαναχρησιμοποιείται συχνά ή κώδικα που είναι περίπλοκος για να συμπεριληφθεί στα ίδια τα πρότυπα, να σημειωθεί ότι αυτοί οι βοηθοί σε καμία περίπτωση δεν πρέπει να εκτελούν βαριά ερωτήματα η να ενημερώνουν τα δεδομένα. Αν όμως αντί για έναν browser υπήρχε μια αίτηση μέσω API GET Request , για να μείνουμε και στο παράδειγμά μας , το αναγνωρίζει και κατευθύνει το αίτημα στην αντίστοιχη λειτουργία ενέργειας του ckan.logic.action.get. Στην περίπτωση αυτή, η λειτουργία ενέργειας θα μπορούσε να είναι package_show, η οποία είναι υπεύθυνη για την ανάκτηση των λεπτομερειών ενός πακέτου δεδομένων. Μόλις το αίτημα προωθηθεί στη λειτουργία package_show από το ckan.logic.action.get, η **Λογική (Logic)** του συστήματος ενεργοποιείται για να εκτελέσει την απαραίτητη επεξεργασία. Εδώ γίνεται ο έλεγχος των δικαιωμάτων του χρήστη για να διασφαλιστεί ότι έχει την απαιτούμενη εξουσιοδότηση για τα δεδομένα αυτά. Τα δεδομένα ανακτώνται μέσω των **Μοντέλων (Models)** του CKAN, που αποτελούν τη δομή της βάσης δεδομένων. Τα μοντέλα οργανώνουν και διαχειρίζονται τα δεδομένα σε δομές που είναι εύκολα διαχειρίσιμες και ανακτήσιμες. Η διαδικασία ολοκληρώνεται με την επιστροφή των δεδομένων στον χρήστη σε μορφή JSON, προσφέροντας μια συνεκτική, κατανοητή και δομημένη απάντηση.



Εικόνα 4.3 Αρχιτεκτονική CKAN

4.2.2 CKAN COMPONENTS

Το CKAN έχει αναπτυχθεί κυρίως με τη Python, γεγονός που το καθιστά ιδιαίτερα προσαρμόσιμο σε διάφορες ανάγκες που απαιτούνται. Η πλατφόρμα βασίζεται σε διάφορες βασικές υπηρεσίες και στοιχεία για να λειτουργεί αποτελεσματικά. Χρησιμοποιεί την **PostgreSQL** για την βάση δεδομένων, για την αποθήκευση μεταδεδομένων σχετικά με σύνολα δεδομένων και άλλες σχετικές πληροφορίες. Επιπλέον, το CKAN ενσωματώνεται συχνά με το **Solr** για δυνατότητες αναζήτησης, ενισχύοντας την ταχύτητα και τη συνάφεια των ερωτημάτων αναζήτησης. Για την εξυπηρέτηση της εφαρμογής CKAN χρησιμοποιούνται συνήθως διακομιστές ιστού όπως ο **Apache** ή ο **Nginx**, ενώ το **Redis** χρησιμοποιείται για την προσωρινή αποθήκευση δεδομένων και την επεξεργασία εργασιών στο παρασκήνιο, γεγονός που βελτιώνει την απόδοση της πλατφόρμας.

Η αρχιτεκτονική του CKAN επιτρέπει την προσθήκη διαφόρων **Plugins** για την επέκταση της λειτουργικότητάς του, ενώ η εξάρτησή του από την Python σημαίνει ότι επωφελείται από τις εκτεταμένες βιβλιοθήκες της Python. Ο συνδυασμός αυτών των τεχνολογιών εξασφαλίζει ότι το CKAN είναι μια ισχυρή και ευέλικτη λύση για τη διαχείριση και δημοσίευση ανοικτών δεδομένων.

Επίσης το API δράσης του CKAN επιτρέπει στους προγραμματιστές να έχουν προγραμματιστική πρόσβαση σε όλες τις βασικές λειτουργίες του CKAN. Επιτρέπει εργασίες όπως η δημιουργία συνόλων δεδομένων, η διαχείριση πόρων και η εκτέλεση αναζητήσεων, καθιστώντας το απαραίτητο για την αυτοματοποίηση και την ενσωμάτωση με άλλα συστήματα.

4.2.3 CKAN EXTENSIONS - PLUGINS

Το CKAN χρησιμοποιεί **plugins** για να επεκτείνει και να προσαρμόσει τη λειτουργικότητά του, τα **plugins** επιτρέπουν την προσαρμογή του CKAN ώστε να ανταποκρίνεται σε συγκεκριμένες ανάγκες χωρίς να μεταβάλλεται η βάση του κώδικα. Το σύστημα των **plugins** στο CKAN υποστηρίζει μια μπορούν προστεθούν, να αφαιρεθούν ή να αντικατασταθούν ανεξάρτητα.

Τα **plugins** που χρειαζόμαστε συνήθως για μια εγκατάσταση CKAN για ένα ολοκληρωμένο σύστημα συλλογής δεδομένων (portal Open Data) είναι τα :

- i. **DataStore** - Χρησιμοποιείται για τη δυνατότητα ισχυρής αναζήτησης και διαχείρισης δεδομένων
- ii. **Datapusher** - απαραίτητο για ενημερώσεις δεδομένων σε πραγματικό χρόνο και συγχρονισμό με το **DataStore**.
- iii. **Recline View** - Δημοφιλής για την παροχή πλούσιων εργαλείων όψεων και εξερεύνησης δεδομένων.
- iv. **Harvest** - προσθέτει λειτουργικότητα για τη συγκομιδή συνόλων δεδομένων από άλλες πηγές, ενισχύοντας τις δυνατότητες εισαγωγής δεδομένων του CKAN.
- v. **DCAT** – Υλοποιεί το πρότυπο **DCAT**, δια λειτουργικότητας μεταδεδομένων ζωτικής σημασίας για την κοινή χρήση και την ολοκλήρωση δεδομένων.

Για να ρυθμίσετε τα πρόσθετα, πρέπει να επεξεργαστείτε το αρχείο `ckan.ini`, αφού έχετε κάνει της απαραίτητες ενέργειες με βάση το **Documentation** του CKAN, το οποίο αρχείο `ckan.ini` συνήθως βρίσκεται στον κατάλογο εγκατάστασης του CKAN. Μέσα σε αυτό το αρχείο, θα βρείτε στο `[app:main]` και προσθέστε ή τροποποιήστε τη ρύθμιση `plugins` για να συμπεριλάβετε τα ονόματα των **plugins** που θέλετε να ενεργοποιήσετε.

```
## DATASTORE && DATAPUSH #####

ckan.datastore.write_url = postgresql://ckan_default:password@localhost/datastore_default

ckan.datastore.read_url = postgresql://datastore_default: password @localhost/datastore_default

ckan.datapusher.url=http://127.0.0.1:8800

ckan.datapusher.api_token=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJYdXhWcEM4N0g1dzVpenFLMkotX09jNlh4dEd6ZldKNlFlQ0d5dGhWYThZliwiaWF0IjoxNzE3OTcyMTE5fQ.QE30gvR_G-hb76VdDnHklsCd07tiXzRiiqW1J7Gte2A

Plugins Settings #####

ckan.plugins = activity datastore datapusher recline_view recline_grid_view recline_graph_view recline_map_view webpage_view
harvest ckan_harvester dcat dcat_rdf_harvester dcat_json_harvester dcat_json_interface structured_data
```

4.2.4 CKAN API

Το Action API του CKAN είναι ένα ισχυρό API τύπου RPC που εκθέτει όλα τα βασικά χαρακτηριστικά του CKAN CLIENT API. Όλες οι βασικές λειτουργίες ενός CKAN (όλα όσα μπορείτε να κάνετε με τη διεπαφή ιστού και άλλα) μπορούν να χρησιμοποιηθούν από εξωτερικό κώδικα που καλεί το API CKAN. Για παράδειγμα, χρησιμοποιώντας το CKAN API η εφαρμογή σας μπορεί:

- Λήψη (Get) JSON αρχεία λιστών με σύνολα δεδομένων, ομάδες ή αλλά αντικείμενα CKAN
- Λήψη μιας πλήρους αναπαράστασης JSON ενός συνόλου δεδομένων, ενός πόρου ή άλλου αντικειμένου
- Αναζήτηση για πακέτα (packages) or δεδομένα (resources) σε ερωτήματα
- Δημιουργία, ενημέρωση και διαγραφή (create,update,delete) συνόλων δεδομένων, πόρων και άλλων αντικειμένων
- Λήψη (GET) μιας ροής δραστηριότητας των πρόσφατα αλλαγμένων συνόλων δεδομένων σε έναν ισότοπο

Για να καλέσετε το API του CKAN, στείλτε ένα αίτημα HTTP POST σε JSON μορφή σε μία από τις διευθύνσεις URL του API του CKAN.

Το CKAN θα επιστρέψει επίσης την απάντησή του σε ένα λεξικό JSON. Ένας τρόπος για να στείλετε ένα λεξικό JSON σε μια διεύθυνση URL είναι η χρήση του πελάτη γραμμής εντολών Curl ή μέσω Postman για την λήψη Δεδομένων , εισαγωγή νέων Data στα υπάρχοντα Packages και Resources η και δημιουργία νέων αρχείων και την εισαγωγή τους στο CKAN μέσω τον API

4.2.5 CKAN – DATASTORE - FILESTORE

Σε αυτήν την ξεχωριστή υπό-ενότητά θα αναφερθούμε σε ένα από τα πιο σημαντικά και δυνατά εργαλεία του CKAN που πραγματοποιείτε η αποθήκευση των δεδομένων και δίνετε η δυνατότητα διαχείρισης τους. Μία από τις βασικές αρχιτεκτονικές επιλογές του CKAN είναι η χρήση δύο ξεχωριστών βάσεων δεδομένων: της κύριας βάσης δεδομένων του CKAN (CKAN Database) και της βάσης δεδομένων του DataStore (DataStore Database). Όμως σημαντικό ρόλο για την αποθήκευση και διαχείριση των αρχείων αποτελεί το **FileStore** που είναι ένας ξεχωριστός χώρος αποθήκευσης και χρησιμοποιείται για την αποθήκευση αρχείων που συνδέονται με τα σετ δεδομένων, όπως αρχεία CSV, JSON, ή ακόμη και άλλα αρχεία.

Η κύρια βάση δεδομένων του CKAN είναι η ckan database μια βάση PostgreSQL που χρησιμοποιείται κύριος για την αποθήκευση των μεταδεδομένων των δεδομένων. Αυτά περιλαμβάνουν πληροφορίες όπως τίτλοι, περιγραφές, ετικέτες, κατηγορίες, πληροφορίες για τον εκδότη, δικαιώματα πρόσβασης και άλλες σχετικές πληροφορίες που βοηθούν στην οργάνωση και την αναζήτηση των δεδομένων. Η CKAN Database διαχειρίζεται επίσης τις πληροφορίες των χρηστών, των οργανισμών και των ρόλων τους εντός της πλατφόρμας.

Από την άλλη η βάση δεδομένων του **DataStore** είναι επίσης μια βάση PostgreSQL και εξειδικεύεται στην αποθήκευση των ίδιων των δεδομένων που περιλαμβάνονται στα σετ δεδομένων του CKAN. Σε αντίθεση με την CKAN Database που αποθηκεύει μεταδεδομένα, το DataStore επιτρέπει την αποθήκευση και την πρόσβαση σε δομημένα δεδομένα μέσω ενός web API. Αυτό επιτρέπει στους χρήστες να πραγματοποιούν σύνθετα ερωτήματα στα δεδομένα, όπως φιλτράρισμα, ταξινόμηση και ανάλυση, χωρίς να χρειάζεται να κατεβάσουν ολόκληρα αρχεία.

Σε αντίθεση με το Datastore το **Filestore**, το οποίο αποθηκεύει ολόκληρα αρχεία χωρίς να παρέχει δυνατότητα πρόσβασης ή αναζήτησης μεμονωμένων τμημάτων του αρχείου, για παράδειγμα φανταστείτε ένα αρχείο CSV ή JSON, στο Filestore αυτό το αρχείο θα αποθηκευτεί ως ολόκληρο αρχείο δηλαδή ως ένα ενιαίο σύνολο στο οποίο δεν θα μπορούμε να έχουμε πρόσβασή σε συγκεκριμένα μέρη μόνο του αρχείου αυτού, όπως βλέπουμε και στην παρακάτω εικόνα, είναι ένα μεγάλο αρχείο CSV από το παράδειγμά μας που θα αναλύσουμε στην παρακάτω ενότητα το οποίο περιέχει της μετρήσεις ποιότητας νερού στις περιοχές της Θεσσαλονίκης.

Dataset Groups Activity Stream Manage

Water Quality Throughout Thessaloniki

Water Quality for Areas : "40 Εκκλησιές" - "Άγιος Ιωάννης" - "Άγιος Παντελεήμονας" - "Άγιος Παύλος" - "Άνω Πόλη" - "Άνω Τούμπα" - "Αμπελόκηποι" - "Ανάληψη" - "Ανθόκηποι" - "Αρετσού" - "ΒΙΠΕΘ" - "Βούλγαρη" - "Βότση" - "ΔΕΘ-ΧΑΝΘ" - "Δενδροτόταμος" - "Διαλογή" - "Ευκαρπία" - "Εύοσμος" - "Ηλιούπολη" - "Κάτω Τούμπα" - "Κέντρο πόλης" - "Καλαμαριά" - "Καραμπουρνάκι" - "Κηφισιά" - "Κορδελιό" - "Κωνσταντινουπολίτικα" - "Μενεμένη" - "Μετέωρα" - "Νέα Κρήνη" - "Νέα Παραλία" - "Νεάπολη" - "Νικόπολη" - "Ντεπώ" - "Ξηροκρήνη" - "Παναγία Φανερωμένη" - "Πανόραμα" - "Πεύκα" - "Πλατεία Δημοκρατίας" - "Πολίχνη" - "Πυλαία" - "Πυλαία (IKEA)" - "Σταυρούπολη" - "Συκίες" - "Σφαγεία" - "Σχολή Τυφλών" - "Τριανδρία" - "Φοίνικας" - "Χαριάσου" - "Ωραιόκαστρο"

Αναζήτηση (π.χ. 2024; Νοέμβριος ή 2024 40 Εκκλησιές)

Data and Resources

CSV Water Quality Throughout Thessaloniki Explore

Προηγούμενο Page 1 of 1 Επόμενο

Εικόνα 4.4 Εικόνα Από το Δικό μου Παράδειγμα της EYAΘ

Σκεφτείτε λοιπόν ότι ίσως να θέλατε να δείτε της μετρήσεις ως πούμε μόνο για την περιοχή του Αγίου Παύλου Θεσσαλονίκης, αυτό μας το επιλύει το Datastore αποθηκεύοντας τα δεδομένα σε έναν πίνακα οπου έχουμε την εύκολη πρόσβαση σε επιλεγμένα δεδομένα που μας αφορούν χωρίς να χρειαστεί να έχουμε περιττή πληροφορία που δεν χρειαζόμαστε είτε ακόμα και με GET request API.

Source: water quality | throughout | thessaloniki

Water Quality Throughout Thessaloniki α;

Fullscreen

Embed

Add Filter

Grid Graph Map 195 records « 1 — 0 »

Q Άγιος Παύλος Go » Filters

_id	Year	Month	Category	Type	Parameter	Value
97458	2024	Ιούνιος	Άγιος Πα...	Φυσικοχ...	Θολότητ...	0,064
97459	2024	Ιούνιος	Άγιος Πα...	Φυσικοχ...	Χρώμα	<5
97460	2024	Ιούνιος	Άγιος Πα...	Φυσικοχ...	Αργίλιο	37
97461	2024	Ιούνιος	Άγιος Πα...	Φυσικοχ...	Χλωριούχα	ΔΠ
97462	2024	Ιούνιος	Άγιος Πα...	Φυσικοχ...	Αγωγιμό...	375
97463	2024	Ιούνιος	Άγιος Πα...	Φυσικοχ...	Συγκέντρ...	7,7
97464	2024	Ιούνιος	Άγιος Πα...	Φυσικοχ...	Υπολειμ...	0,25
97465	2024	Ιούνιος	Άγιος Πα...	Μικροβιο...	Κολοβακ...	ΔΑ
97466	2024	Ιούνιος	Άγιος Πα...	Μικροβιο...	Escheric...	ΔΑ
97467	2024	Ιούνιος	Άγιος Πα...	Μικροβιο...	Enteroco...	ΔΑ
97468	2024	Ιούνιος	Άγιος Πα...	Μικροβιο...	Αριθμός ...	<1
97469	2024	Ιούνιος	Άγιος Πα...	Οργανολ...	Οσμή	αποδεκτή
97470	2024	Ιούνιος	Άγιος Πα...	Οργανολ...	Γεύση	αποδεκτή
91635	2024	Μάιος	Άγιος Πα...	Φυσικοχ...	Θολότητ...	0,082
91636	2024	Μάιος	Άγιος Πα...	Φυσικοχ...	Χρώμα	<5
91637	2024	Μάιος	Άγιος Πα...	Φυσικοχ...	Αργίλιο	<25

Εικόνα 4.5 Εικόνα Με τα αποτελέσματα από το Water Quality (δικό μου παράδειγμα) της ΕΥΑΘ

```

1. GET http://127.0.0.1:5000/api/3/action/datastore_search
2. {   "resource_id": "9dd8393f-9728-40a1-96f6-168241e953ae",
3.     "limit": 100000000,
4.     "filters": {
5.         "Year": "2023",
6.         "Category": "Άγιος Παύλος"
7.     }
8. }

```

Γενικά μπορούμε να πούμε ότι η αρχιτεκτονική αυτή προσφέρει μια ευέλικτη και αποδοτική λύση για την διαχείριση των ανοιχτών δεδομένων καθώς επιτρέπει την βελτιστοποίηση των λειτουργιών αυτών την αύξηση της απόδοσης την βελτίωση της ασφάλειας καθώς και την εύκολη επέκταση καθιστώντας το ιδανικό για την υποστήριξη μεγάλων συνόλων δεδομένων σε διάφορους τομείς.

4.3 Data Catalog Vocabulary (DCAT)

Η κοινή χρήση πόρων δεδομένων μεταξύ διαφορετικών οργανισμών, ερευνητών, κυβερνήσεων και πολιτών απαιτεί την παροχή μεταδεδομένων. Το DCAT (Data Catalog Vocabulary) είναι ένα πρότυπο RDF που ενισχύει τη δια λειτουργικότητα μεταξύ των καταλόγων δεδομένων βασίζονται στο διαδίκτυο, που αναπτύχθηκε από το W3C. Παρέχει ένα ενιαίο μοντέλο και λεξιλόγιο για την περιγραφή συνόλων δεδομένων και υπηρεσιών δεδομένων, το οποίο ενισχύει τη συγκέντρωση μεταδεδομένων και την ευρεσιμότητα σε πολλαπλούς καταλόγους, το οποίο αναπτύχθηκε αρχικά στο πλαίσιο κυβερνητικών καταλόγων δεδομένων, όπως και , αλλά είναι επίσης εφαρμόσιμο και έχει χρησιμοποιηθεί και σε άλλα πλαίσια. Με τη χρήση του DCAT, οι εκδότες μπορούν να δημιουργήσουν αποκεντρωμένους καταλόγους και να επιτρέψουν ομοσπονδιακές αναζητήσεις σε διαφορετικούς ισότοπους με έναν συνεπή μηχανισμό ερωτημάτων. Συνολικά, το DCAT διευκολύνει έναν πιο συνδεδεμένο και

αναζητήσιμο ιστό δεδομένων, βελτιώνοντας την πρόσβαση και την ενσωμάτωση σε διαφορετικές πηγές δεδομένων.

Το DCAT παρέχει κλάσεις και ιδιότητες RDF για να επιτρέπει την περιγραφή και τη συμπερίληψη συνόλων δεδομένων και υπηρεσιών δεδομένων σε έναν κατάλογο. Η χρήση ενός τυποποιημένου μοντέλου και λεξιλογίου διευκολύνει την κατανάλωση και τη συγκέντρωση μεταδεδομένων από πολλαπλούς καταλόγους, οι οποίοι μπορούν:

1. να αυξήσει την ευρεσιμότητα των συνόλων δεδομένων και των υπηρεσιών δεδομένων
2. να επιτρέπει την ομοσπονδιακή αναζήτηση συνόλων δεδομένων σε καταλόγους σε πολλαπλές τοποθεσίες

Τα δεδομένα που περιγράφονται σε έναν κατάλογο μπορούν να έχουν πολλές μορφές, από υπολογιστικά φύλλα, μέσω XML και RDF έως διάφορες εξειδικευμένες μορφές. Το DCAT δεν κάνει καμία υπόθεση σχετικά με αυτές τις μορφές των συνόλων δεδομένων, αλλά κάνει διάκριση μεταξύ του αφηρημένου συνόλου δεδομένων και των διαφόρων εκδηλώσεων ή διανομών του.

4.3.1 DCAT-AP

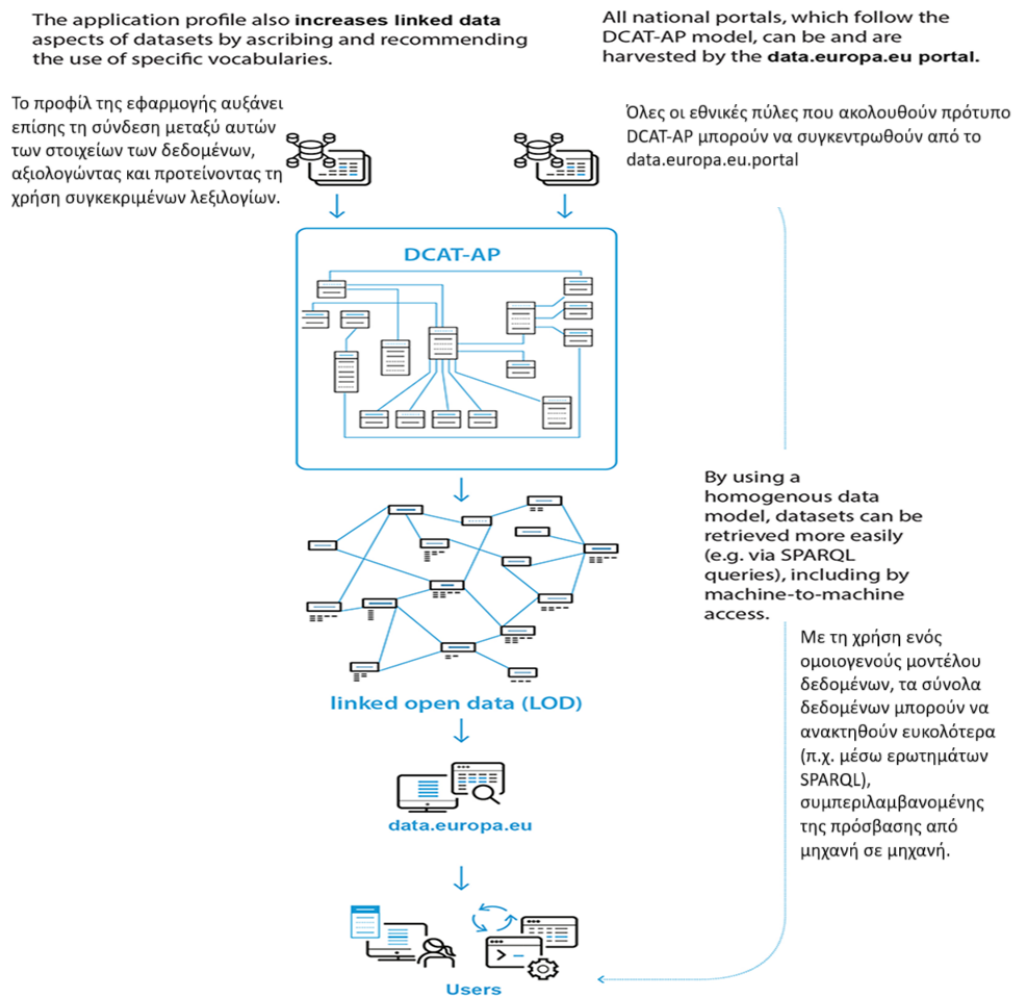
Βασισμένο στο λεξιλόγιο καταλόγου δεδομένων (DCAT), το DCAT-AP είναι μια προδιαγραφή μεταδεδομένων που έχει σχεδιαστεί για να βελτιώσει την ευρεσιμότητα και τη διαλειτουργικότητα των συνόλων δεδομένων του δημόσιου τομέα σε ολόκληρη την Ευρώπη, διευκολύνει την αναζήτηση και την ανταλλαγή δεδομένων μεταξύ των πυλών, καθιστώντας τα δεδομένα του δημόσιου τομέα πιο προσιτά.

Η Ευρωπαϊκή Επιτροπή, στην ανακοίνωσή της για τα ανοικτά δεδομένα το 2011, τόνισε τη σημασία των αναγνώσιμων από μηχανήματα μεταδεδομένων θα διευκολύνει τη διαλειτουργικότητα και αυξάνει την αξία των δεδομένων για επαναχρησιμοποίηση. Το DCAT-AP συμβάλλει στην ικανοποίηση αυτών των αναγκών παρέχοντας έναν δομημένο τρόπο περιγραφής συνόλων δεδομένων, διευκολύνοντας τους χρήστες να εντοπίζουν και να κατανοούν τα δεδομένα που είναι διαθέσιμα σε διαφορετικές πύλες.

Με την τυποποίηση του τρόπου περιγραφής των μεταδεδομένων και τη χρήση ελεγχόμενων λεξιλογίων όπως το EuroVoc και άλλων προτύπων όπως το Dublin Core και το INSPIRE, το DCAT-AP διασφαλίζει την ομαλή συνεργασία διαφορετικών συστημάτων δεδομένων.

Ένα από τα κύρια οφέλη της DCAT-AP είναι η εξασφάλιση της αξιοπιστίας, παρέχοντας ένα πρότυπο για την περιγραφή των μεταδεδομένων που δημοσιεύονται από πύλες δεδομένων σε όλη την Ευρώπη βοηθώντας τους χρήστες να επαναχρησιμοποιήσουν τα δεδομένα επιτρέποντας και την πολύγλωσση πρόσβαση αν τα δεδομένα βρίσκονται σε άλλο κράτος μέλος έτσι η επαναχρησιμοποίηση των δεδομένων αυξάνετε σημαντικά καθιστώντας τα δεδομένα αναζητήσιμα και προσβάσιμα με την καταχώρισή τους σε μία ή περισσότερες πύλες δεδομένων, γεγονός που μπορεί να μειώσει σημαντικά το κόστος.

Η χρήση του DCAT-AP εξασφαλίζει ότι τα σύνολα δεδομένων περιγράφονται με εναρμονισμένο τρόπο σύμφωνα με συγκεκριμένα λεξιλόγια. Αυτό επιτρέπει τη σύνδεση των συνόλων δεδομένων με σύνολα δεδομένων από άλλες πηγές που περιγράφονται επίσης με τη χρήση αυτού του μοντέλου δεδομένων. Με τη χρήση ενός ομογενούς μοντέλου δεδομένων, τα σύνολα δεδομένων μπορούν να ανακτηθούν πιο εύκολα (π.χ. μέσω ερωτημάτων SPARQL), συμπεριλαμβανομένης της πρόσβασης από μηχανή σε μηχανή.



Εικόνα 4.6 Περιγραφή Μοντέλου DCAT-AP Ευρωπαϊκή Επιτροπή

4.3.2 CKAN – DCAT

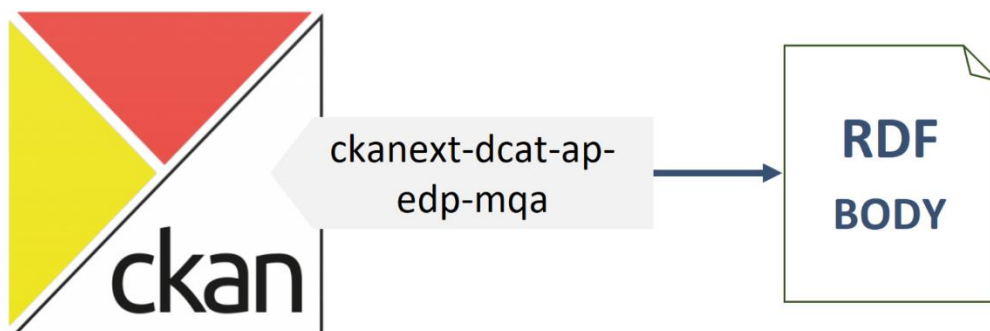
Με την άνοδο των παγκόσμιων πρωτοβουλιών για τα ανοικτά δεδομένα, η ανάγκη για τυποποιημένη ανταλλαγή μεταδεδομένων μεταξύ διαφορετικών καταλόγων γίνεται ολοένα και πιο σαφής. Πύλες Ανοικτών Δεδομένων της ΕΕ συγκεντρώνουν σύνολα δεδομένων από διάφορες πηγές, αναδεικνύοντας την απαίτηση για μια σαφή και τυποποιημένη διεπαφή που θα επιτρέπει την απρόσκοπτη ενσωμάτωση μεταδεδομένων. Υπάρχει αυξανόμενη συμφωνία ότι το DCAT (Data Catalog Vocabulary) αποτελεί βασική λύση για το σκοπό αυτό

Στο CKAN, αυτή η επέκταση εισάγει διάφορα χαρακτηριστικά: Εσωτερικά, αυτά τα χαρακτηριστικά υλοποιούνται χρησιμοποιώντας: Αυτές οι εξελίξεις διευκολύνουν την καλύτερη ενσωμάτωση και κοινή χρήση μεταδεδομένων, απλοποιώντας τη διαδικασία σύνδεσης συνόλων δεδομένων σε διαφορετικές πλατφόρμες.

- ο [RDF DCAT EndPoint](#) που εκθέτουν τα σύνολα δεδομένων του καταλόγου σε διαφορετικά RDF (decat plugin).

- ο [RDF Harvester](#) που επιτρέπει την εισαγωγή RDF από άλλους καταλόγους για τη δημιουργία συνόλων δεδομένων CKAN (dcat_rdf_harvester plugin).
- ο [JSON DCAT Harvester](#) που επιτρέπει την εισαγωγή αντικειμένων JSON που βασίζονται σε όρους DCAT αλλά δεν έχουν οριστεί ως JSON-LD

Το CKAN μπορεί επίσης να λειτουργήσει με το DCAT-AP μέσω Harvest, υποστηρίζει την ενσωμάτωση και συγχρονισμό μεταδεδομένων από άλλες πύλες ανοιχτών δεδομένων και, με την επέκταση DCAT, μπορεί να συλλέξει σύνολα δεδομένων που περιγράφονται με το πρότυπο DCAT-AP.



Εικόνα 4.7 CKAN-DCAT

Για να ενεργοποιηθούν αυτές οι λειτουργίες, θα πρέπει να ενημερωθεί το αρχείο `ckan.ini` για τη κατάλληλη διαμόρφωση και των πρόσθετων στα plugins.

```
1. ## CKAN DCAT #####
2. ckanext.dcat.enable_rdf_endpoints = True
3. ckanext.dcat.catalog_endpoint = /dcat/catalog.{_format}
4. ckanext.dcat.dataset_endpoint = https://{ckan-instance-host}/dataset/{dataset-id}.{format}
5. ckanext.dcat.base_uri = http://127.0.0.1:5000
6. ckanext.dcat.publisher_uri = http://example.com/organization/your-org
7. ckanext.dcat.publisher_name = Aggelos Eyath
8. ckanext.dcat.publisher_email = info@example.com
9. ckanext.dcat.rdf.profiles = euro_dcat_ap
10. ckanext.dcat.content_negotiation = true
```

4.4 Επίλογος

Οι πύλες ανοιχτών δεδομένων διευκολύνουν την ελεύθερη διακίνηση των πληροφοριών, ενισχύοντας πολλαπλούς τομείς. Σύμφωνα με τα παραπάνω καταλήγουμε στο συμπέρασμα ότι η σημασία αλλά και οι δυνατότητες των ανοιχτών δεδομένων είναι πολλαπλές και ιδιαίτερα ωφέλιμες. Παράλληλα είναι σαφές ότι τα δεδομένα από μόνα τους δεν μπορούν να επιφέρουν αλλαγές καθώς είναι απαραίτητες οι πλατφόρμες και οι πύλες που χρησιμοποιούνται για τη διάδοση και τη διαχείρισή τους. Πιο συγκεκριμένα με τη συμβολή των πυλών αυτών το κοινό αποκτά ελεύθερη πρόσβαση σε μια ευρεία γκάμα δεδομένων τα οποία μπορεί να τα χρησιμοποιεί κατά το δοκούν με αποτέλεσμα τόσο το κοινό όσο και οι δημόσιοι φορείς να διευκολύνονται. Η χρήση αυτή τους επιτρέπει να προβούν στην εφαρμογή νέων τεχνολογιών, να μπορούν να μελετούν τα δεδομένα για στατιστικούς και ερευνητικούς λόγους και να επεκτείνουν τις δυνατότητες κάθε δημόσιου φορέα με απώτερο σκοπό το γενικό καλό. Επιπροσθέτως

η συμβολή των διαφόρων πλατφορμών διαχείρισης δεδομένων καθιστά πιο εύκολη τη διαχείριση του μεγάλου όγκο δεδομένων που εισέρχονται καθημερινά σε ένα σύστημα ή μια βάση δεδομένων. Όπως αναφέρθηκε και παραπάνω πλατφόρμες όπως το CKAN επιδρούν ουσιαστικά στον τρόπο που μπορούν να επεξεργαστούν τα δεδομένα αλλά και στο πως μπορούν να αποθηκευτούν έτσι ώστε σε δεύτερο χρόνο να είναι εύκολη η ανάκτηση τους για παράδειγμα με λέξεις κλειδιά και να μπορούν να είναι αναγνώσιμα από τους χρήστες. Η χρησιμότητα κάθε πλατφόρμας είναι διαφορετική καθώς παρέχουν ένα οργανωμένο περιβάλλον για την αποθήκευση, τη συντήρηση και την αναζήτηση δεδομένων. Η ποικιλία που προσφέρετε με βάση τη δομή κάθε πλατφόρμας ή πύλης ανοικτών δεδομένων έχει ως αποτέλεσμα ο καθένας να αποκτά πρόσβαση σε δεδομένα χρήσιμα για διάφορες κατηγορίες που απευθύνονται τόσο για τους πολίτες και επαγγελματιών όσο και σε δημόσιους φορείς. Στο σημείο αυτό να επισημάνουμε την σημασία της δια λειτουργικότητας και της τυποποίησης καθώς οι πύλες ανοικτών δεδομένων πρέπει να τεθούν σε κάποιο πλαίσιο ώστε να είναι λειτουργικές και να μπορεί να χρησιμοποιηθεί το περιεχόμενο τους αποτελεσματικά. Η χρήση προτύπων, όπως το DCAT, ενισχύει τη δια λειτουργικότητα σε σχέση με τις διαφορετικές κατηγορίες δεδομένων. Η πλατφόρμα επιτρέπει στους οργανισμούς να δημοσιεύουν δεδομένα, να τα συνοδεύουν με μεταδεδομένα και να παρέχουν στους χρήστες τη δυνατότητα να τα αναζητούν καθώς και να τα επεξεργάζονται. Κλείνοντας να επισημάνουμε έναν βασικό παράγοντα, αυτόν της επεκτασιμότητας, όπου με την δυνατότητα προσθήκης διαφόρων λογισμικών, τα λεγόμενα plugins, αλλά και με τη χρήση διαφόρων τεχνολογιών όπως το API του CKAN, αυξάνονται οι παροχές που είναι σε θέση να προσφέρει η κάθε πλατφόρμα. Επομένως με την συμβολή των επεκτάσεων αυτών αυξάνονται οι δυνατότητες κάθε πλατφόρμας κάνοντας την εμπειρία των χρηστών όλο και καλύτερη και πιο αποτελεσματική συμβάλλοντας στη βελτίωση πολλαπλών τομέων στην καθημερινότητα.

Κεφάλαιο 5ο: Παραδείγματα με Δεδομένα της ΕΥΑΘ

5.1 ΕΙΣΑΓΩΓΗ

Η εργασία αυτή επικεντρώνεται στη δημιουργία και διαχείριση μιας διαδικτυακής πλατφόρμας ανοικτών δεδομένων για την ποιότητα του νερού στην περιοχή της Θεσσαλονίκης, χρησιμοποιώντας την πλατφόρμα CKAN. Το CKAN είναι ένα ισχυρό σύστημα διαχείρισης και διανομής ανοικτών δεδομένων που χρησιμοποιείται ευρέως από μεγάλους οργανισμούς παγκοσμίως, λόγω της ευκολίας προσαρμογής του και της δυνατότητας προσαρμογής του στις ανάγκες κάθε οργανισμού. Για τους σκοπούς του παρόντος έργου, τα δεδομένα που αφορούν την ποιότητα του νερού συλλέχθηκαν από την ΕΥΑΘ (την εταιρεία ύδρευσης) και αποθηκεύτηκαν σε τρία διαφορετικά σύνολα δεδομένων

Σύνολο δεδομένων 1: Μαζική Οργάνωση δεδομένων για την ποιότητα του νερού

Το πρώτο σύνολο δεδομένων περιέχει πληροφορίες σχετικά με την ποιότητα του νερού, αποθηκευμένες σε έναν πίνακα με τις ακόλουθες στήλες:

- **_id:** Ένα μοναδικό αναγνωριστικό για κάθε καταχώρηση.
- **Έτος:** Το έτος καταγραφής.
- **Μήνας:** Ο μήνας της καταγραφής.
- **Κατηγορία (περιοχή):** Η περιοχή στην οποία έγινε η μέτρηση.
- **Τύπος:** Το είδος της εξέτασης (π.χ. «Φυσικοχημική», «Μικροβιολογική», «Οργανοληπτική»).
- **Παράμετρος:** Οι συγκεκριμένες παράμετροι που μετρήθηκαν για κάθε τύπο.
- **Τιμή:** Η καταγεγραμμένη τιμή για κάθε παράμετρο.

Σε αυτό το σύνολο δεδομένων, οι στήλες «Τύπος» και «Παράμετρος» συνδυάζονται μοναδικά. Συγκεκριμένα, κάθε τύπος εξέτασης περιλαμβάνει συγκεκριμένες παραμέτρους, ως εξής:

- ❖ **Φυσικοχημικές παράμετροι:**
 - Θολρότητα NTU
 - Χρώμα
 - Αλουμίνιο
 - Χλωρίδια
 - Αγωγιμότητα
 - Συγκέντρωση ιόντων υδρογόνου
 - Υπολειμματικό χλώριο
- ❖ **Μικροβιολογικές παράμετροι:**
 - Κολοβακτηρίδια
 - Escherichia coli (E. coli)
 - Εντερόκοκκοι
 - Αριθμός αποικιών στους 22°C
- ❖ **Οργανοληπτικές παράμετροι:**
 - Οσμή
 - Γεύση

Σύνολο δεδομένων 2: Ανά τύπο και περιοχή

Το δεύτερο σύνολο δεδομένων οργανώνεται με βάση τον τύπο της εξέτασης και την περιοχή. Οι χρήστες μπορούν να έχουν πρόσβαση σε τρία διαφορετικά αρχεία CSV, καθένα από τα οποία περιέχει δεδομένα για μια συγκεκριμένη περιοχή και τύπο εξέτασης. Για κάθε τύπο, υπάρχουν καταγεγραμμένες τιμές παραμέτρων για όλες τις διαθέσιμες ημερομηνίες καταγραφής. Για παράδειγμα:

Πίνακας για τις οργανοληπτικές εξετάσεις:

ID | ΕΤΟΣ | ΜΗΝΑΣ | ΠΕΡΙΟΧΗ | ΤΥΠΟΣ | ΠΑΡΑΜΕΤΡΟΣ | ΤΙΜΗ

Παράμετροι (Οσμή, γεύση)

Πίνακας για φυσικοχημικές εξετάσεις:

ID | ΕΤΟΣ | ΜΗΝΑΣ | ΠΕΡΙΟΧΗ | ΤΥΠΟΣ | ΠΑΡΑΜΕΤΡΟΣ | ΤΙΜΗ

Παράμετροι (Θολότητα NTU, Χρώμα, Αλουμίνιο, Χλωρίδια, Αγωγιμότητα, Συγκέντρωση ιόντων υδρογόνου)

Πίνακας για τις μικροβιολογικές εξετάσεις:

ID | ΕΤΟΣ | ΜΗΝΑΣ | ΠΕΡΙΟΧΗ | ΤΥΠΟΣ | ΠΑΡΑΜΕΤΡΟΣ | ΤΙΜΗ

Παράμετροι (Κολοβακτηριοειδή, *Escherichia coli* (*E. coli*), Εντερόκοκκοι, Αριθμός αποικιών στους 22°C)



Εικόνα 5.1 Water Quality Group By Area (Thessaloniki) & Type DATASET

Σύνολο δεδομένων 3: Ανάλυση ανά χρονική περίοδο τύπο παραμέτρου και περιοχή

Το τρίτο σύνολο δεδομένων είναι οργανωμένο ανά έτος, μήνα, περιοχή και τύπο παραμέτρου. Αυτό το σύνολο δεδομένων περιλαμβάνει τον μεγαλύτερο όγκο δεδομένων, καθώς ενημερώνεται συνεχώς με νέες εγγραφές κάθε μήνα. Με αυτόν τον τρόπο, οι χρήστες μπορούν να αναζητήσουν πληροφορίες με βάση μια συγκεκριμένη χρονική περίοδο και περιοχή, όπως βλέπουμε και στην παρακάτω εικόνα

Data and Resources		
	2024_Μάιος_40_Εκκλησιές_Φυσικοχημικές	Explore
	2024_Μάιος_40_Εκκλησιές_Μικροβιολογικές	Explore
	2024_Μάιος_40_Εκκλησιές_Οργανοληπτικές	Explore
	2024_Απρίλιος_40_Εκκλησιές_Φυσικοχημικές	Explore
	2024_Απρίλιος_40_Εκκλησιές_Μικροβιολογικές	Explore
	2024_Απρίλιος_40_Εκκλησιές_Οργανοληπτικές	Explore
	2024_Φεβρουάριος_40_Εκκλησιές_Φυσικοχημικές	Explore
	2024_Φεβρουάριος_40_Εκκλησιές_Μικροβιολογικές	Explore
	2024_Φεβρουάριος_40_Εκκλησιές_Οργανοληπτικές	Explore

Εικόνα 5.2 Water Quality Group By Area (Thessaloniki) , Type & Date Dataset

Το συγκεκριμένο portal που έχουμε στήσει CKAN επιτρέπει στους χρήστες να αναζητούν, να φιλτράρουν και να μεταφορτώνουν δεδομένα σε διάφορες μορφές αρχείων, όπως CSV και JSON. Επιπλέον, παρέχει δυνατότητες διαδραστικής απεικόνισης δεδομένων, επιτρέποντας στους χρήστες να εξάγουν πληροφορίες σχετικά με την ποιότητα του νερού είτε μαζικά είτε ανά συγκεκριμένη περιοχή και χρονική περίοδο όπως αυτοί επιθυμούν καθώς επιπλέον όλα τα datasets έχουν άδεια χρήσης Open Data Commons Public Domain Dedication and License (PDDL)

3 datasets found

Order by: Relevance ▼

Licenses: [Open Data Commons Public Domain Dedication and License \(PDDL\) x](#)

Water Quality Group By Area (Thessaloniki) & Type

Water Quality for Areas : "40 Εκκλησιές" - "Άγιος Ιωάννης" - "Άγιος Παντελεήμονας" - "Άγιος Παύλος" - "Άνω Πόλη" - "Άνω Τούμπα" - "Αμπελόκηποι" - "Ανάληψη" - "Ανθόκηποι" -...

CSV

Water Quality Group By Area (Thessaloniki) , Type & Date

Area , Type & Date (Περιοχή),(Τυπος),(Ημέρα)

CSV

Water Quality Throughout Thessaloniki

Water Quality for Areas : "40 Εκκλησιές" - "Άγιος Ιωάννης" - "Άγιος Παντελεήμονας" - "Άγιος Παύλος" - "Άνω Πόλη" - "Άνω Τούμπα" - "Αμπελόκηποι" - "Ανάληψη" - "Ανθόκηποι" -...

CSV

Εικόνα 5.3 Άδεια Χρήσης των Datasets

Η δημιουργία αυτής της πλατφόρμας έχει την προοπτική στην αύξηση της διαφάνειας όσον αφορά τα δεδομένα για την ποιότητα του νερού, καθιστώντας τα προσβάσιμα τόσο στους ειδικούς όσο και στο ευρύ κοινό. Τέτοια εργαλεία ενισχύουν την ευαισθητοποίηση του κοινού σε θέματα περιβαλλοντικής υγείας και προωθούν την ανοικτή πρόσβαση σε δεδομένα.

Η χρήση της πλατφόρμας CKAN ανέδειξε τις δυνατότητές της για την προώθηση των ανοικτών δεδομένων και επιβεβαίωσε τη σημασία των σύγχρονων τεχνολογικών εργαλείων για τη διαχείριση και την παροχή πληροφοριών που σχετίζονται με το περιβάλλον.

5.2 Γιατί τα Δεδομένα της ΕΥΑΘ ;

Επιλέξαμε τα δεδομένα της ΕΥΑΘ (Εταιρεία Ύδρευσης & Αποχέτευσης Θεσσαλονίκης) για τη συγκεκριμένη εργασία, καθώς αφορούν άμεσα την ποιότητα του νερού που καταναλώνεται στην περιοχή της Θεσσαλονίκης, κάτι που αποτελεί κρίσιμη πληροφορία για τη δημόσια υγεία και το περιβάλλον. Τα δεδομένα αυτά είναι ιδιαίτερα σημαντικά για ειδικούς, ερευνητές, αλλά και για το ευρύ κοινό, καθώς παρέχουν πληροφόρηση σχετικά με τις φυσικοχημικές, μικροβιολογικές και οργανοληπτικές παραμέτρους του νερού.



Εικόνα 5.4 Λογότυπο της ΕΥΑΘ

Ωστόσο, παρατηρήσαμε ότι τα δεδομένα της ΕΥΑΘ δεν είναι εύκολα διαθέσιμα ή προσβάσιμα με τρόπο που να διευκολύνει την αξιοποίησή τους από τους χρήστες και ειδικούς. Παρότι υπάρχουν κάποιες πληροφορίες διαθέσιμες στις ιστοσελίδες, δεν υπάρχουν οργανωμένα και εύκολα προσβάσιμα αρχεία, με αποτέλεσμα να είναι δύσκολο να βρεθούν και να αξιοποιηθούν από τις μηχανές αναζήτησης ή πλατφόρμες που ειδικεύονται στη συλλογή και διαχείριση δεδομένων. Αυτό σημαίνει ότι τα δεδομένα που θα μπορούσαν να είναι εξαιρετικά χρήσιμα για ερευνητικούς σκοπούς ή την ανάλυση της ποιότητας του νερού παραμένουν δυσπρόσιτα, περιορίζοντας την πραγματική τους αξία.

Επιπλέον, δεν υπάρχει διαθέσιμο API (Application Programming Interface) για τα δεδομένα αυτά, κάτι που θα μπορούσε να διευκολύνει την αυτοματοποιημένη πρόσβαση και χρήση τους από προγραμματιστές και ερευνητές. Ένα API θα επέτρεπε σε ειδικούς να ενσωματώνουν τα δεδομένα σε εφαρμογές, να κάνουν αναλύσεις σε πραγματικό χρόνο και να δημιουργούν πιο σύνθετα εργαλεία για τη μελέτη της ποιότητας του νερού. Χωρίς API, η χρήση των δεδομένων γίνεται περισσότερο περίπλοκη και απαιτεί επιπλέον χειροκίνητη εργασία.

5.3 Λειτουργίες και API που χρησιμοποιήσαμε

Στην ενότητα αυτή παρουσιάζεται η ανάπτυξη και υλοποίηση του συστήματος διαχείρισης δεδομένων ποιότητας νερού και η απόκτηση των δεδομένων και η αποθήκευσή αυτών, αξιοποιώντας τις λειτουργίες της πλατφόρμας του CKAN. Το CKAN, λόγω του ανοιχτού κώδικα που διαθέτει, το

καθιστά εξαιρετικά ευέλικτο και παραμετροποιήσιμο στην εκάστοτε ανάγκη που μπορεί να προκύψει καθώς σου επιτρέπει όχι μόνο τη χρήση των ενσωματωμένων λειτουργιών του αλλά και την ανάπτυξη εξειδικευμένων λύσεων μέσω custom υλοποιήσεων.

Πιο συγκεκριμένα στραφήκαμε στις δυνατότητες του API (EndPoint) του CKAN το οποίο σου επιτρέπει πολύ εύκολα την εισαγωγή πληροφορίας στα είδη υπάρχοντα Datasets η και την δημιουργία των datasets μέσω του API καθώς βεβαία και την εξαγωγή αυτών. Η υποδομή του CKAN όπως είδαμε και προηγουμένως είναι δομημένη στην γλώσσα προγραμματισμού αυτής της Python σε συνδυασμό με την βάση δεδομένων της Postgress , δύο ισχυρές τεχνολογίες που διασφαλίζουν τη σταθερότητα, την επεκτασιμότητα και την αξιόπιστη διαχείριση δεδομένων.

Για τέλος η παραμετροποίηση του CKAN συμπληρώνεται με custom κώδικα, παρέχοντας λύσεις προσαρμοσμένες στις ανάγκες του έργου, όπως η ενσωμάτωση δεδομένων μέσω web scraping ή με συνδυασμό το API και τη χρήση stored procedures στην Βάση δεδομένων για αποδοτική επεξεργασία και ανάλυση της πληροφορίας αξιοποιώντας την στο μέγιστο.

Όσο αφορά το Front -End της πλατφόρμας παρέχει ένα ισχυρό σύνολο εργαλείων και μεθόδων για την παραμετροποίηση και τη διαχείριση του, επιτρέπει την πλήρη τροποποίηση της εμφάνισης και της λειτουργικότητας, προσαρμόζοντας την πλατφόρμα στις εκάστοτε ανάγκες του χρήστη. Έτσι λοιπόν το CKAN χρησιμοποιεί ένα σύνολο σύγχρονων τεχνολογιών που το καθιστούν ευέλικτο και εύκολα παραμετροποιήσιμο για το front-end. Μέσω των **Jinja2 templates**, του **Bootstrap framework** και εργαλείων όπως το **Webassets** και το **htmx**, προσφέρει δυνατότητες για την ανάπτυξη responsive διασαφών και δυναμικών στοιχείων. Επιπλέον, με τη χρήση **CSS**, **JavaScript** και του **CKAN API**, δίνεται η δυνατότητα διαχείρισης δεδομένων και εμφάνισης resources σε πραγματικό χρόνο, Επιπλέον υποστήριξη helper functions και jQuery plugins διευκολύνει την ανάπτυξη λειτουργιών, επιτρέποντας την προσαρμογή του CKAN στις ανάγκες κάθε έργου, ενώ παράλληλα διασφαλίζεται ένα φιλικό και λειτουργικό περιβάλλον για τον τελικό χρήστη.

5.3.1 Λειτουργίες Διαχειρίσεις Δεδομένων (BACK-END) API

Η λειτουργικότητα που εφαρμόσαμε για την ανάκτηση δεδομένων και την προσθήκη των Νέων δεδομένων είναι κατά κυρίως λόγος με το API του CKAN και διαφορές procedure και function στην Βάση Δεδομένων του CKAN και DATASTORE ανάλογα. Δημιουργήσαμε λοιπόν το Python Αρχείο

WaterQuality_To_Datastore.py, όπου εκεί μέσα έχουμε τις κλάσεις WaterQualityScrape, CkanRequest, ExecuteSql και την Main όπου εκτελώντας την, το πρόγραμμα πραγματοποιεί όλα τα απαραίτητα βήματα για να συλλέξει τα νέα δεδομένα (scraping), να τα συγκρίνει με όσα ήδη είναι αποθηκευμένα, και, τέλος, να τα εισάγει στο CKAN, DataStore, εκτελώντας επιπλέον τις σχετικές διαδικασίες (Procedures) και συναρτήσεις (Functions) στις δυο αυτές βάσεις δεδομένων.

Πιο συγκεκριμένα:

Η κλάση **WaterQualityScrape** είναι υπεύθυνη για την ανάκτηση δεδομένων (scraping) από ιστοσελίδα της ΕΥΑΘ που παρέχονται οι μετρήσεις ποιότητας νερού στις περιοχές της Θεσσαλονίκης. Αρχικά, η μέθοδος *scrape_categories* διαβάζει την αρχική σελίδα και εντοπίζει όλες τις «κατηγορίες» / περιοχές. Έπειτα, για κάθε περιοχή, η *scrape_month_year* εντοπίζει όλες τις ημερομηνίες που υπάρχουν ανά μήνα και έτος. Τέλος, για κάθε σύνδεσμο που αφορά έναν συγκεκριμένο μήνα/έτος/περιοχή, η *scrape_table* διαβάζει τους πίνακες της σελίδας και εξάγει τις πραγματικές τιμές των μετρήσεων (τύπος, παράμετρος, τιμή). Έτσι, η κλάση συγκεντρώνει όλα τα απαραίτητα στοιχεία από τον ιστότοπο, έτοιμα για επεξεργασία ή εισαγωγή σε κάποια βάση δεδομένων.

Προχωρώντας λοιπόν στην επομένη κλάση έχουμε ορίσει να είναι η κλάση που θα αναλαμβάνει την επικοινωνία με το CKAN API ώστε τα δεδομένα που συγκεντρώθηκαν να ελεγχθούν, να εισαχθούν ή να δημιουργήσουν καινούργια DataStore resources. Στην ουσία, παρέχει τις μεθόδους **Get_data_all** που γίνεται η ανάκτηση όλων των εγγραφών από το κύριο Resource που είναι συγκεντρωμένα όλα τα data σε έναν πίνακα με την βοήθεια της μεθόδου GET όπως :

```
1. GET http://127.0.0.1:5000/api/3/action/datastore_search?resource_id=9dd8393f-9728-40a1-96f6-168241e953ae
```

οπού μάς επιστρέφει σε μορφή JSON τα στοιχεία του πίνακα με Resource ID το οποίο έχει ορισθεί στο παραπάνω URL. Στη συνέχεια, για την εισαγωγή των νέων δεδομένων, έπειτα από ελέγχους που γίνονται με τη βοήθεια των **functions insert_new_data** και **create_datastore**, εισάγονται τα νέα δεδομένα του μήνα στους πίνακες Περιοχή και Τύπο Παραμέτρου (αναφερθήκαμε στην παραπάνω ενότητα), χρησιμοποιώντας τη μέθοδο POST.

Request μεθόδου Insert New Data για την εισαγωγή νέων δεδομένων σε ένα συγκεκριμένο Resource ID με την μέθοδο POST του CKAN API.

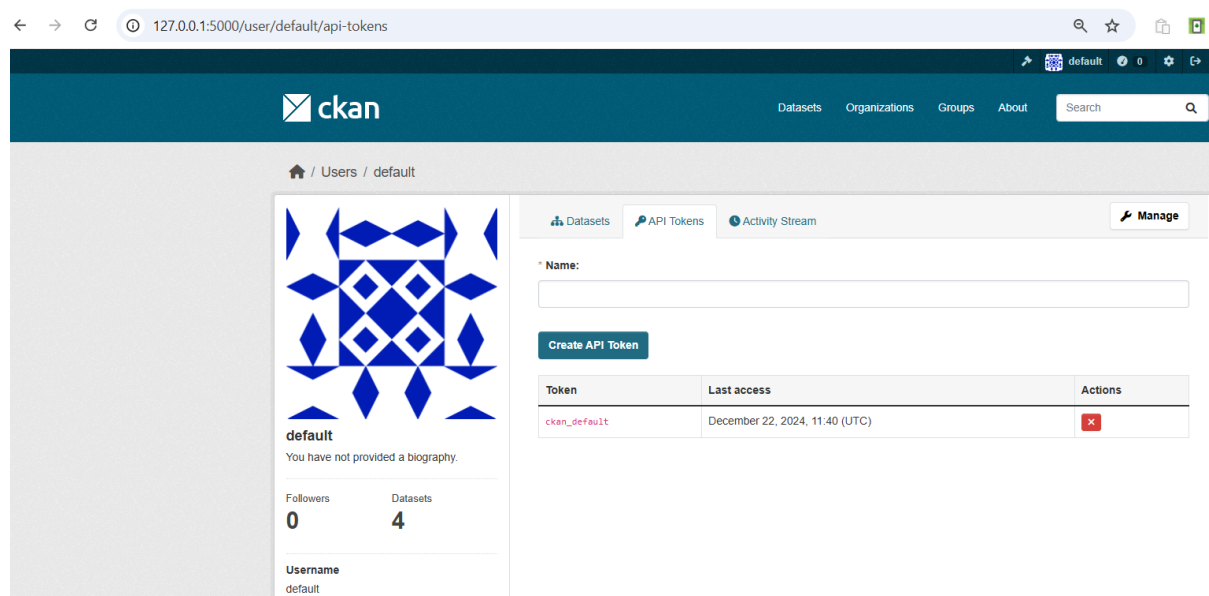
```
1. def insert_new_data(api_key, resource_id, new_data):
2.     datastore_data = {
3.         'resource_id': resource_id,
4.         'records': new_data,
5.         'method': 'insert',
6.         "force": "True"
7.     }
8.
9.     try:
10.         response = requests.post(
11.             f'http://127.0.0.1:5000/api/3/action/datastore_upsert',
12.             headers={'X-CKAN-API-Key': api_key},
13.             json=datastore_data
14.         ).....
```

Request μεθόδου Create datastore για την δημιουργία νέων Resources σε συγκεκριμένο Package χρησιμοποιώντας την μέθοδο POST του CKAN API

```
1. def create_datastore(api_key, fields, records):
2.     year_month_records = {}
3.
4.     for record in records:
5.         year = str(record["Year"]).replace(" ", "_")
6.         month = record["Month"].replace(" ", "_")
7.         category = record["Category"].replace(" ", "_")
8.         type_ = record["Type"].replace(" ", "_")
9.         key = f"{year}_{month}_{category}_{type_}"
10.
11.         if key not in year_month_records:
12.             year_month_records[key] = []
13.
14.         year_month_records[key].append(record)
15.     create_url = 'http://127.0.0.1:5000/api/3/action/datastore_create'
16.     headers = {'Content-Type': 'application/json', 'X-CKAN-API-Key': api_key}
17.     for key, records in year_month_records.items():
18.         resource = {
19.             "resource": {
20.                 "package_id": 'e24f0295-06aa-410a-8087-e6aaa5011a70',
21.                 "name": key,
22.                 "datastore_name": key,
23.                 "datastore_active": True,
24.                 "format": "CSV" },
25.             "fields": fields,
26.             "records": records,
27.             "method": "insert",
28.             "force": True } .....
```

Να σημειώσουμε ότι για την δημιουργία νέων Datastores και για την εισαγωγή νέων δεδομένων σε αυτών είναι απαραίτητο να διαθέτουμε ένα έγκυρο **CKAN_API_KEY**. Το κλειδί αυτό λειτουργεί ως μέσο αυθεντικοποίησης, επιτρέποντας στους χρήστες να πραγματοποιούν ασφαλείς ενέργειες στο CKAN. Χωρίς το CKAN_API_KEY, δεν υπάρχει η απαιτούμενη πρόσβαση για να δημιουργήσουν νέα Datastores ή να εισάγουν νέα δεδομένα, διασφαλίζοντας έτσι την προστασία και την ακεραιότητα των πληροφοριών. Το API_KEY μπορούμε να το παράξουμε μέσα από τον λογαριασμό του CKAN κάνοντας create API token, όπως φαίνεται και στην παρακάτω εικόνα και το εισάγουμε οπου είναι απαραίτητο ακόμα και στην παραμετροποίηση του Datapusher στο config αρχείο του CKAN.

```
1. ckan.datapusher.url=http://127.0.0.1:8800
2. ckan.datapusher.api_token=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJYdXhWcEM4N0g1dzVpenFLMkotX09jNlh4dEd6ZldKNlFIQ0d5dGhWYThZiIiwiaWF0IjoxNzE3OTcyMTExfQ.QE30gvR_G-hb76VdDnHklsCd07tiXzRiiqW1J7Gte2A
3.
```



Εικόνα 5.5 Δημιουργία API TOKEN

Για παράδειγμα, στο παρακάτω κομμάτι γίνεται μια νέα τυχαία εισαγωγή του έτους 2026, μήνα Ιανουάριο, με κατηγορία (Περιοχή) sampleCategory, Τύπο Α και παράμετρο X με την τιμή 123.45.

```
POST--
http://127.0.0.1:5000/api/3/action/datastore_upsert?Authorization=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1NiJ9.eyJqdGkiOiJYdXhWcEM4N0g1dzVpenFLMkotX09jNlh4dEd6ZldKNlFIQ0d5dGhWYThZiIiwiaWF0IjoxNzE3OTcyMTExfQ.QE30gvR_G-hb76VdDnHklsCd07tiXzRiiqW1J7Gte2A
```

```
1. { "resource_id": "d3e94bc4-648b-443c-8d99-0913017877d2",
2.   "records": [{
3.     "Year": "2026", // Updated value
4.     "Month": "January",
5.     "Category": "SampleCategory",
6.     "Type": "TypeA",
7.     "Parameter": "ParameterX",
8.     "Value": "123.45"}]
10.  "method": "insert",
11.  "force": true}
```

Έπειτα δημιουργούνται νέα αρχεία CSV με την μέθοδο **create_datastore** για την δημιουργία CSV αρχείων ανά Περιοχή Τύπο και Ημερομηνία(Μήνα/Ετος),καθώς έχουμε διαθέσιμο το API από το CKAN ,απαιτείται και η εισαγωγή του `package_id` για τον ορισμό του resource, ώστε να ανήκει κατευθείαν σε κάποιο package. Για παράδειγμα, με το παρακάτω, δημιουργούμε ένα Resource στο συγκεκριμένο `package_id` με format CSV, κάνοντας την εισαγωγή των γραμμών στις οριζόμενες στήλες

```

1. POST -- http://127.0.0.1:5000/api/3/action/datastore_create
2. {   "resource": {
3.     "package_id": "e24f0295-06aa-410a-8087-e6aaa5011a70",
4.     "name": "January_2026_Category_ParameterX",
5.     "datastore_name": "January_2026_Category_ParameterX",
6.     "datastore_active": true,
7.     "format": "CSV"
8.   },
9.   "fields": [
10.    { "id": "Year", "type": "text" },
11.    { "id": "Month", "type": "text" },
12.    { "id": "Category", "type": "text" },
13.    { "id": "Type", "type": "text" },
14.    { "id": "Parameter", "type": "text" },
15.    { "id": "Value", "type": "text" }
16.  ],
17.  "records": [
18.    {
19.      "Year": "2026",
20.      "Month": "January",
21.      "Category": "SomeCategory",
22.      "Type": "ParameterX",
23.      "Parameter": "pH",
24.      "Value": "7.0"
25.    },
26.    {
27.      "Year": "2026",
28.      "Month": "January",
29.      "Category": "SomeCategory",
30.      "Type": "ParameterX",
31.      "Parameter": "Turbidity",
32.      "Value": "0.5"
33.    }
34.  ],
35.  "method": "insert",
36.  "force": true}

```

Σε συνδυασμό λοιπόν με τις παραπάνω διαδικασίες ολοκληρώνουμε την λειτουργία με την εκτέλεση SQL procedures και Function κυρίως για την εισαγωγή των δεδομένων στην βάση του Datastore υπολογίζοντας ποια δεδομένα και σε πύον πίνακα του Datastore θα αποθηκευτούν τα δεδομένα καθώς κάθε Resource ID είναι και πίνακας στην βάση δεδομένων του Datastore έτσι ώστε να μπορεί ο χρήστης ευκολά να διαβάζει τα δεδομένα χωρίς απαραίτητο να κατεβάσει το αρχείο CSV που καθιστά μια εύκολη πρόσβαση στον καθένα, αυτό το πετυχαίνει η εκτέλεση της procedure `al_procedureinserttoresources()` που εκτελείται από την βάση του Datastore,η Procedure `al_updateTimeToMainDatastore()` ενημερώνει την τελευταία ώρα εκτέλεσής και η function `al_insertViews()` κάνει εισαγωγή των νέων resources που δημιουργήσαμε στον πίνακα `resource_view` οπου είναι απαραίτητος για την διαχείριση/εμφάνιση των δεδομένων.

Συνοψίζοντας, η **main** μέθοδος συγκεντρώνει όλες τις παραπάνω λειτουργίες και εκτελεί τις διαδικασίες: καλεί πρώτα την κλάση **WaterQualityScrape** για την ανάκτηση των στοιχείων, στη συνέχεια τη **CkanRequest** για τη διαχείριση αυτών των δεδομένων στο CKAN (ενημέρωση, εισαγωγή, δημιουργία νέων CSV resources) και, τέλος, ενεργοποιεί τις **SQL procedures** και **functions** στη βάση, ώστε να εισαχθούν ορθά οι εγγραφές. Έτσι, με την εκτέλεση του κεντρικού αυτού τμήματος,

επιτυγχάνεται η πλήρης ροή από την ανάκτηση δεδομένων μέχρι την οργάνωση και εμφάνισή τους στο CKAN DataStore

Όσον αφορά την παραπάνω διαδικασία, δημιουργήθηκε ένα **custom plugin**, το ckanext-WaterThessaloniki, εκτελώντας την εντολή **ckan generate extension** όπως περιγράφεται στην τεκμηρίωση του CKAN (ReadTheDocs, παράγραφος 5.1.2 "Creating a new extension"). Η εντολή αυτή δημιουργεί τη βασική δομή του extension, όπως τον φάκελο ckanext-WaterThessaloniki, στο path /home/aggelos/ckan/lib/default/src. Κατά την εκτέλεση της εντολής, συμπληρώθηκαν τα παρακάτω στοιχεία:

1. (default) aggelos@DESKTOP:~/ckan/lib/default/src/ ckan generate extension
2. Extension's name [must begin 'ckanext-']: ckanext-WaterThessalonik
3. Author's name []: Aggelos
4. Author's email []: aggelos@gmail.com
5. Your Github user or organization name []: aggelos@github.com
6. Brief description of the project []:
7. List of keywords (separated by spaces) [CKAN]:
8. Do you want to include code examples? [y/N]: n

Το συγκεκριμένο extension διευκολύνει την ενημέρωση των δεδομένων ή τη δημιουργία νέων συνόλων δεδομένων, όπως περιγράφεται παραπάνω, για χρήση μόνο από τους διαχειριστές (Admin). Το plugin αυτό εκτελεί το αρχείο WaterQuality_To_Datastore.py μέσω του κουμπιού «Update Datasets», το οποίο διαχειρίζεται την αυτόματη εισαγωγή ή ενημέρωση των δεδομένων στο CKAN DataStore και παρέχει κατάλληλα μηνύματα για την επιτυχή ολοκλήρωση ή τυχόν σφάλματα, εξασφαλίζοντας την αποτελεσματικότητα στη διαχείριση των δεδομένων.

Παρακάτω παρατίθεται απόσπασμα του κώδικα που χρησιμοποιείται στο αρχείο **views.py**, το οποίο δημιουργήθηκε αυτόματα κατά την εκτέλεση της εντολής ckan generate extension. Το απόσπασμα διαχειρίζεται την εκτέλεση του αρχείου **WaterQuality_To_Datastore.py**:

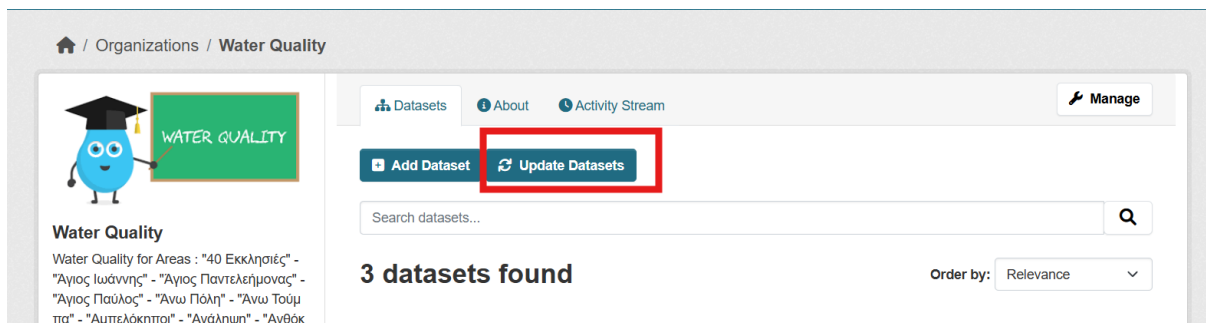
```

1. script_path = "/home/aggelospincha/ckan/lib/default/src/ckan/ckan/CustomScripts/WaterQuality_To_DataStore.py"
2. watertessaloniki.route("/start", methods=["POST"])
3. def start_script():
4.     global status_info
5.     with lock:
6.         if status_info["status"] == "running":
7.             return jsonify({"message": "Script is already running."}), 400
8.
9.     command = f". /usr/lib/ckan/default/bin/activate && python3 {script_path}"
10.    process = subprocess.Popen(
11.        command,
12.        stdout=subprocess.PIPE,
13.        stderr=subprocess.PIPE,
14.        shell=True,
15.        preexec_fn=os.setsid
16.    ) ... Συνεχίζεται

```

Για τη σωστή λειτουργία του plugin, είναι απαραίτητο να δηλωθεί στο **ckan.plugins** στο αρχείο διαμόρφωσης ckan.in. Επίσης, απαιτείται η κατάλληλη παραμετροποίηση στο αρχείο **plugin.py**, όπου καθορίζεται η βασική λειτουργικότητα του extension και οι μέθοδοι που ενεργοποιούνται από το CKAN

Επιπλέον, τυχόν τροποποιήσεις ή προσθήκες απαιτούν την επεξεργασία των αρχείων που βρίσκονται στον φάκελο του εκάστοτε extension στην προκειμένη περίπτωση στο /ckanext-WaterThessaloniki.



Εικόνα 5.6 Update Datasets (ενημερώνει τα δεδομένα)

Η ολοκληρωμένη αυτή διαδικασία, που συνδυάζει Python (για scraping και CKAN API) και PostgreSQL (για την αποθήκευση και την εκτέλεση procedures/functions), παρέχει έναν μηχανισμό άμεσης, αυτοματοποιημένης και ενημέρωσης των δεδομένων. Έτσι, οι αρμόδιοι ή οι τελικοί χρήστες έχουν πάντα πρόσβαση σε φρέσκα και δομημένα στοιχεία, δίχως να χρειάζεται να τα διαχειρίζονται χειροκίνητα.

5.3.2 Λειτουργία Εύκολης Αναζήτησης

Για να διευκολυνθεί η εύρεση συγκεκριμένων δεδομένων σε μια λίστα αποτελεσμάτων, ενσωματώθηκε μια λειτουργία αναζήτησης (search bar) ώστε ο χρήστης να μπορεί να εισάγει λέξεις-κλειδιά ή φράσεις (π.χ. έτος, περιοχή) και να φιλτράρει άμεσα τα διαθέσιμα αρχεία. Επιπλέον, επειδή ο συνολικός αριθμός των αρχείων είναι μεγάλος, επιλέχθηκε να εμφανίζονται το πολύ δέκα αποτελέσματα σε κάθε σελίδα, διασφαλίζοντας ότι η περιήγηση στα δεδομένα παραμένει γρήγορη και κατανοητή.

Στην ενότητα αυτή παρουσιάζεται η ανάπτυξη και υλοποίηση του συστήματος διαχείρισης δεδομένων ποιότητας νερού και η απόκτηση των δεδομένων και η αποθήκευσή αυτών, αξιοποιώντας τις λειτουργίες της πλατφόρμας του CKAN. Το CKAN, λόγω του ανοιχτού κώδικα που διαθέτει, το καθιστά εξαιρετικά ευέλικτο και παραμετροποιήσιμο στην εκάστοτε ανάγκη που μπορεί να προκύψει καθώς σου επιτρέπει όχι μόνο τη χρήση των ενσωματωμένων λειτουργιών του αλλά και την ανάπτυξη εξειδικευμένων λύσεων μέσω custom υλοποιήσεων.

Πιο συγκεκριμένα στραφήκαμε στις δυνατότητες του API (EndPoint) του CKAN το οποίο σου επιτρέπει πολύ εύκολα την εισαγωγή πληροφορίας στα είδη υπάρχοντα Datasets η και την δημιουργία των datasets μέσω του API καθώς βεβαία και την εξαγωγή αυτών. Η υποδομή του CKAN όπως είδαμε και προηγουμένως είναι δομημένη στην γλώσσα προγραμματισμού αυτής της Python σε συνδυασμό με την βάση δεδομένων της Postgress, δύο ισχυρές τεχνολογίες που διασφαλίζουν τη σταθερότητα, την επεκτασιμότητα και την αξιόπιστη διαχείριση δεδομένων.

Όσο αφορά το Front -End της πλατφόρμας παρέχει ένα ισχυρό σύνολο εργαλείων και μεθόδων για την παραμετροποίηση και τη διαχείριση του, επιτρέπει την πλήρη τροποποίηση της εμφάνισης και της λειτουργικότητας, προσαρμόζοντας την πλατφόρμα στις εκάστοτε ανάγκες του χρήστη. Έτσι λοιπόν το CKAN χρησιμοποιεί ένα σύνολο σύγχρονων τεχνολογιών που το καθιστούν ευέλικτο και εύκολα παραμετροποιήσιμο για το front-end. Μέσω των **Jinja2 templates**, του **Bootstrap framework** και

εργαλείων όπως το **Webassets** και το **htmx**, προσφέρει δυνατότητες για την ανάπτυξη responsive διασαφών και δυναμικών στοιχείων. Επιπλέον, με τη χρήση **CSS**, **JavaScript** και του **CKAN API**, δίνεται η δυνατότητα διαχείρισης δεδομένων και εμφάνισης resources σε πραγματικό χρόνο, Επιπλέον υποστήριξη helper functions και jQuery plugins διευκολύνει την ανάπτυξη λειτουργιών, επιτρέποντας την προσαρμογή του CKAN στις ανάγκες κάθε έργου, ενώ παράλληλα διασφαλίζεται ένα φιλικό και λειτουργικό περιβάλλον για τον τελικό χρήστη.

The screenshot displays the CKAN interface for the dataset 'Water Quality Group By Area (Thessaloniki), Type & Date'. The sidebar on the left shows the organization 'Water Quality' with a logo of a blue character holding a sign that says 'WATER QUALITY'. The main content area features a search bar with the text '2023;40;φω'. Below the search bar, there is a section titled 'Data and Resources' listing four CSV files: '2023_Δεκέμβριος_40_Εκκλησιές_Φυσικοχημικές', '2023_Νοέμβριος_40_Εκκλησιές_Φυσικοχημικές', '2023_Οκτώβριος_40_Εκκλησιές_Φυσικοχημικές', and '2023_Σεπτέμβριος_40_Εκκλησιές_Φυσικοχημικές'. Each file has an 'Explore' button. At the bottom, there is a pagination section with 'Previous', 'Page 1 of 1', and 'Next' buttons, and a 'Go' button. Below the pagination, there are tags for 'Microbiological parameters', 'Organoleptic parameters', 'Physico-chemical parameters', 'Thessaloniki', and 'water'.

Εικόνα 5.6 Μπάρα Αναζήτησης

Για την υλοποίηση αυτών των δυνατοτήτων, αναπτύξαμε κώδικα JavaScript στο αρχείο **read.html**. Ο JavaScript κώδικας διαχειρίζεται τη λειτουργικότητα της αναζήτησης και του φιλτραρίσματος των αποτελεσμάτων με βάση τις εισαγόμενες λέξεις-κλειδιά ή φράσεις. Επιπλέον, ο κώδικας ελέγχει και ενημερώνει τη σελιδοποίηση, εξασφαλίζοντας ότι μόνο δέκα αποτελέσματα εμφανίζονται ανά σελίδα και ότι οι χρήστες μπορούν να προηγηθούν εύκολα μεταξύ των σελίδων. Αυτή η δυναμική προσέγγιση όχι μόνο βελτιώνει την απόδοση της εφαρμογής, αλλά παρέχει επίσης μια φιλική προς τον χρήστη εμπειρία, καθιστώντας την αναζήτηση και την περιήγηση στα δεδομένα πιο αποτελεσματική και ευχάριστη.

5.3.3 DCAT-RDF - LINKED DATA (Παράδειγμα με τα δεδομένα της ΕΥΑΘ)

Σε αυτήν την παράγραφο θα εξετάσουμε το RDF-DCAT, αξιολογώντας το κατά πόσο πληροί τις προϋποθέσεις για **Linked Data**, αν συμμορφώνεται με το **5-Star Linked Data** και αν χρησιμοποιεί σωστά τα **vocabularies**. Για την ενεργοποίηση του DCAT-RDF, προστέθηκαν και εγκαταστάθηκαν τα απαραίτητα Plugins στο αρχείο **ckan.ini** σύμφωνα με το **documentation** του CKAN. Το RDF του Dataset λοιπόν **Water Quality Group By Area (Thessaloniki) & Type** είναι διαθέσιμο στο URL

<http://127.0.0.1:5000/dataset/water-quality-group-by-area-thessaloniki-type.RDF>

Αυτό το URL ξεκινά κατευθείαν τη λήψη του RDF αρχείου. Παρακάτω παρουσιάζεται ένα μικρό απόσπασμα από το αρχείο RDF με ένα dcat:Distribution και ένα παράδειγμα SPARQL ερωτήματος, το οποίο υλοποιείται με τη βοήθεια του Apache Jena Fuseki

```
<? 1. xml version="1.0" encoding="utf-8"?>
2. <rdf:RDF
3.   xmlns:dc="http://www.w3.org/ns/dcat#"
4.   xmlns:dct="http://purl.org/dc/terms/"
5.   xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
6.   xmlns:foaf="http://xmlns.com/foaf/0.1/"
7.   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
8. >
9.<dc:Dataset rdf:about="http://127.0.0.1:5000/dataset/0395beb3-bebd-4f97-b401-a1259fcae298">
10.<dct:title>Water Quality Group By Area (Thessaloniki) & Type</dct:title>
11.<dct:description>Water Quality for Areas :
12. "40 Εκκλησιές" - "Άγιος Ιωάννης" - "Άγιος Παντελεήμονας" - "Άγιος Παύλος"
13. </dct:description>
14.   <dc:landingPage rdf:resource="https://quality.eyath.gr/">
15.   <dct:identifier>0395beb3-bebd-4f97-b401-a1259fcae298</dct:identifier>
16.   <dc:keyword>Microbiological parameters</dc:keyword>
17.   <dc:keyword>Organoleptic parameters</dc:keyword>
18.   <dc:keyword>Physico-chemical parameters</dc:keyword>
19.   <dc:keyword>Thessaloniki</dc:keyword>
20.   <dc:keyword>water</dc:keyword>
21.   <dc:keyword>water Quality</dc:keyword>
22.<dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2024-07-05T21:56:18.721842</dct:issued>
23.   <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2024-09-01T16:07:58.052207</dct:modified>
24.   <dc:contactPoint>
25.     <vcard:Organization rdf:nodeID="Nbdb3f93445e548459109412457e302fb">
26.       <vcard:fn>Aggelos</vcard:fn>
27.       <vcard:hasEmail rdf:resource="mailto:aggelos@gmail.com"/>
28.     </vcard:Organization>
29.   </dc:contactPoint>
30.   <dc:contactPoint>
31.     <vcard:Organization rdf:nodeID="N9d6fc244eef44f90a40e1712a16bf78d">
32.       <vcard:fn>Aggelos</vcard:fn>
33.       <vcard:hasEmail rdf:resource="mailto:aggelos@gmail.com"/>
34.     </vcard:Organization>
35.   </dc:contactPoint>
36.   <dct:publisher>
37.     <foaf:Organization rdf:about="http://127.0.0.1:5000/organization/d39deae5-5781-4b6c-bb1e-e46c74488175">
38.       <foaf:name>Water Quality</foaf:name>
39.     </foaf:Organization>
40.   </dct:publisher>
41.   <dc:distribution>
42.     <dc:Distribution rdf:about="http://127.0.0.1:5000/dataset/0395beb3-bebd-4f97-b401-a1259fcae298/resource/13eba7cd-d816-485b-a525-9a4d8dc7f419">
43.       <dct:title>40 Εκκλησιές_Φυσικοχημικές</dct:title>
44.       <dct:format>CSV</dct:format>
45.       <dc:accessURL rdf:resource="http://127.0.0.1:5000/datastore/dump/13eba7cd-d816-485b-a525-9a4d8dc7f419"/>
46.       <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2024-07-05T21:57:45.473932</dct:issued>
47.       <dct:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2024-07-20T23:05:54.579767</dct:modified>
48.     </dc:Distribution>
49.   </dc:distribution>
```

```

1. PREFIX dcat: <http://www.w3.org/ns/dcat#>
2. PREFIX dct: <http://purl.org/dc/terms/>
3.
4. SELECT *
5. WHERE {
6.   ?distribution a dcat:Distribution ;
7.               dct:title ?title ;
8.               dct:format ?format ;
9.               dcat:accessURL ?accessURL .
10.  FILTER (regex(?title, "^40_Εκκλησιές", "i"))}

```

Table Response 3 results in 0.024 seconds Simple view Ellipse Filter query results Page size: All

distribution	title	format	accessURL
1 <http://127.0.0.1:5000/dataset/0395beb3-bebd-4f97-b401-a1259fcae298/resource/95db1123-0d0b-4b54-8739-25a0a52f534b>	40_Εκκλησιές_Μικροβιο λογικές	CSV	<http://127.0.0.1:5000/datastore/dump/95db1123-0d0b-4b54-8739-25a0a52f534b>
2 <http://127.0.0.1:5000/dataset/0395beb3-bebd-4f97-b401-a1259fcae298/resource/b73ae20e-65f7-492c-a176-f65e6fa76a9e>	40_Εκκλησιές_Οργανολ ηττικές	CSV	<http://127.0.0.1:5000/datastore/dump/b73ae20e-65f7-492c-a176-f65e6fa76a9e>
3 <http://127.0.0.1:5000/dataset/0395beb3-bebd-4f97-b401-a1259fcae298/resource/13eba7cd-d816-485b-a525-9a4d8dc7f419>	40_Εκκλησιές_Φυσικοχη μικές	CSV	<http://127.0.0.1:5000/datastore/dump/13eba7cd-d816-485b-a525-9a4d8dc7f419>

Showing 1 to 3 of 3 entries << < 1 > >>

Εικόνα 5.6 Αποτέλεσμα από το πάνω SPARQL ερώτημα

Αξιολόγηση ως Linked Data

Το RDF-DCAT παράδειγμα χρησιμοποιεί μοναδικά URIs και διαχωρίζει τα URIs από τα περιεχόμενα, επιτρέποντας ακριβείς αναφορές. Ωστόσο, το παράδειγμα πληροί τα πρώτα τέσσερα αστέρια του μοντέλου, προσφέροντας διαθεσιμότητα στο web, μηχανικά αναγνώσιμα δεδομένα, και χρήση διακριτών URIs. Για το 5ο αστέρι όμως, απαιτείται η χρήση δημόσιων URLs και η ενίσχυση των συνδέσεων με άλλες πηγές δεδομένων.

5.3.4 Εργαλεία που Χρησιμοποιήθηκαν

Για την ανάπτυξη και υλοποίηση του έργου χρησιμοποιήθηκαν διάφορα λογισμικά και τεχνολογικά εργαλεία που μας παρείχαν τις απαραίτητες δυνατότητες για την επίτευξη του αποτελέσματος αυτού. Τα εργαλεία αυτά διευκόλυναν την αποτελεσματική διαχείριση του κώδικα, την οργάνωση και διαχείριση των βάσεων δεδομένων καθώς και την τη δοκιμή των API Request

Τα εργαλεία που χρησιμοποιήθηκαν είναι τα εξής :

Visual Studio Code

Χρησιμοποιήθηκε ως κύριο περιβάλλον ανάπτυξης για τον προγραμματισμό σε Python. Παρέχει υποστήριξη για debugging, και επεκτάσεις που διευκολύνουν την ανάπτυξη και διαχείριση του κώδικα.

pgAdmin4

Χρησιμοποιήθηκε ως εργαλείο διαχείρισης της βάσης δεδομένων. Το pgAdmin4 προσφέρει ένα φιλικό περιβάλλον για την εκτέλεση ερωτημάτων SQL και τη διαχείριση των δεδομένων.

Postman

Χρησιμοποιήθηκε για τις δοκιμές των API requests. Το Postman επιτρέπει την εύκολη δημιουργία, αποστολή και ανάλυση αιτημάτων προς τα API, διευκολύνοντας την επαλήθευση της σωστής λειτουργίας τους.

Apache Jena Fuseki

Για τον έλεγχο λειτουργίας RDF και ερωτημάτων SPARQL χρησιμοποιήθηκε το Apache Jena Fuseki οπου είναι SPARQL server

Virtual Machine Ubuntu 22.04.3 LTS

Η εγκατάσταση του CKAN πραγματοποιήθηκε σε μια εικονική μηχανή με λειτουργικό σύστημα Ubuntu 22.04.3 LTS. Η χρήση του Virtual Machine επέτρεψε την απομόνωση του περιβάλλοντος ανάπτυξης και την ευκολότερη διαχείριση των εξαρτήσεων.

5.4 Συμπεράσματα Εργασίας

Στην παρούσα εργασία εξετάσαμε τα **Ανοιχτά Δεδομένα**, αναλύοντας τι είναι, πότε και γιατί πρέπει να αξιοποιηθούν, καθώς και τα οφέλη από την εφαρμογή τους. Αναφερθήκαμε στα Ανοικτά Δημόσια Δεδομένα και τη σημασία τους για την καθημερινή ζωή, τονίζοντας πώς η διαθεσιμότητά τους μπορεί να βελτιώσει τη διαφάνεια, την αποτελεσματικότητα και τη συμμετοχή των πολιτών σε διάφορους τομείς. Ένα κρίσιμο μέρος της εργασίας ήταν η ανάκτηση δεδομένων ποιότητας νερού από τις περιοχές της Θεσσαλονίκης, όπου τα δεδομένα ανακτήθηκαν από την ΕΥΑΘ και οργανώθηκαν σε μια πύλη ανοικτών δεδομένων CKAN. Παρατηρήσαμε ότι, αν και η ΕΥΑΘ διαθέτει τα σημαντικά αυτά δεδομένα για την ποιότητα του νερού, αυτά δεν ήταν εύκολα αναγνωρίσιμα από μηχανές αναζήτησης ή προσβάσιμα σε αυτά ή σε ένα συγκεκριμένο μέρος αυτών. Η μετατροπή αυτών των δεδομένων σε συνδεδεμένα δεδομένα μέσω της πύλης ανοικτών δεδομένων επιτρέπει πλέον την ευκολότερη πρόσβαση, ανάλυση και χρήση από τους ερευνητές και το ευρύ κοινό.

Διαθέτοντας δεδομένα για την ποιότητα των υδάτων σε μια ανοικτή δημόσια πύλη, ανοίγονται πολλές δυνατότητες. Για παράδειγμα, οι πολίτες μπορούν να παρακολουθούν την ποιότητα του νερού στην περιοχή τους σε πραγματικό χρόνο, ενισχύοντας την ευαισθητοποίησή τους και την ικανότητά τους να λαμβάνουν χρήσιμες αποφάσεις για την υγεία τους. Επιπλέον, οι αρμόδιοι μπορούν να αξιοποιούν τα δεδομένα αυτά για την ανάπτυξη στρατηγικών πρόληψης και αντιμετώπισης της περιβαλλοντικής ρύπανσης, προωθώντας ένα πιο βιώσιμο περιβάλλον. Οι επιχειρήσεις στον τομέα της περιβαλλοντικής διαχείρισης μπορούν επίσης να χρησιμοποιήσουν τα δεδομένα για την ανάπτυξη καινοτόμων λύσεων που βελτιώνουν την ποιότητα του νερού.

Η συγκέντρωση δεδομένων σε πύλες ανοικτών δεδομένων έχει πολλαπλά θετικά αποτελέσματα. Καθιστά τα δεδομένα διαθέσιμα σε όλους, προωθεί τη συνεργασία και την ανταλλαγή γνώσεων μεταξύ διαφορετικών τομέων και αυξάνει την εμπιστοσύνη των πολιτών στις δημόσιες υπηρεσίες. Επιπλέον, η ορθή διαχείριση και κοινή χρήση δεδομένων μπορεί να συμβάλει στην πρόληψη πιθανών προβλημάτων, όπως η έγκαιρη ανίχνευση και αντιμετώπιση της ρύπανσης των υδάτων, προστατεύοντας έτσι την υγεία και το περιβάλλον.

Συνοψίζοντας, η παρούσα εργασία ανέδειξε τη σημασία των ανοικτών δεδομένων και τον αντίκτυπό τους στην κοινωνία και το περιβάλλον. Ο μετασχηματισμός των δεδομένων για την ποιότητα των υδάτων σε εύκολα προσβάσιμες μορφές και η καταχώρησή τους σε ανοικτές πλατφόρμες βελτιώνει τη διαφάνεια, προωθεί την ενεργό συμμετοχή των πολιτών και συμβάλλει σε ένα πιο βιώσιμο και υγιές μέλλον.

5.5 Σχέδια Μελλοντικής Εξέλιξης του Συστήματος

Η εργασία αυτή επικεντρώνεται στα ανοιχτά δεδομένα αναλύοντας βασικές έννοιες, αρχές και εργαλεία που τα υποστηρίζουν. Στο μέλλον, θα μπορούσε να επεκταθεί πολύ εύκολα με πιο συγκεκριμένα παραδείγματα εφαρμογής στην πράξη έτσι ώστε να φανεί πώς τα ανοιχτά δεδομένα χρησιμοποιούνται σε πραγματικές συνθήκες. Επιπλέον, θα μπορούσε να γίνει μια απλή έρευνα για το πώς αυτά τα δεδομένα θα μας επηρεάσουν την ζωή στην Ελλάδα. Συγκεντρώνοντας, λοιπόν τα δεδομένα ποιότητας νερού στις περιοχές της Θεσσαλονίκης θα μπορούσε κάλλιστα να επεκταθεί αυτή η συγκέντρωση δεδομένων και σε άλλες περιοχές της Ελλάδας εξετάζοντας τους τρόπους με τους οποίους θα μπορούσε να βελτιωθεί η χρήση των ανοικτών δεδομένων στη χώρα μας, μέσα από την εκπαίδευση των πολιτών ή την καλύτερη οργάνωση αυτών των δεδομένων από δημόσιους φορείς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] 5 Star Open Data, «5 Star Open Data,» [Ηλεκτρονικό]. Available: <https://5stardata.info/en/>.
- [2] Open Government Partnership, «5ο Εθνικό Σχέδιο Δράσης 2023-2025,» ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ - Υπουργείο Ψηφιακής Διακυβέρνησης, 2024.
- [3] Brain Traffic, «An Introduction to Metadata and Taxonomies,» [Ηλεκτρονικό]. Available: <https://www.braintraffic.com/blog/an-introduction-to-metadata-and-taxonomies>.
- [4] European Data Portal, «Benefits and value of open data - data.europa.eu,» [Ηλεκτρονικό]. Available: <https://data.europa.eu/en/publications/datastories/benefits-and-value-open-data>.
- [5] CKAN documentation, «CKAN documentation Release 2.12.0a0 CKAN contributors,» [Ηλεκτρονικό]. Available: <https://readthedocs.org/projects/ckan/downloads/pdf/latest/>.
- [6] CKAN, «CKAN Extensions CKAN + DCAT,» [Ηλεκτρονικό]. Available: <https://extensions.ckan.org/extension/dcat/#overview>.
- [7] W3C, «Data Catalog Vocabulary (DCAT),» 2014. [Ηλεκτρονικό]. Available: <https://travesia.mcu.es/server/api/core/bitstreams/1e9db3df-2e70-4012-8160-adc0ce4a44aa/content>.
- [8] DKAN, «DKAN Open Data Portal,» [Ηλεκτρονικό]. Available: <https://dkan.readthedocs.io/en/7.x-1.x/index.html>. [Πρόσβαση 2024].
- [9] European Data Portal, «European Union Open Data Portal,» [Ηλεκτρονικό]. Available: <https://data.europa.eu/en>.
- [10] geodata, «geodata,» gov.gr, [Ηλεκτρονικό]. Available: <http://www.geodata.gov.gr/>.
- [11] w3.org, «<https://www.w3.org/resources/>,» [Ηλεκτρονικό]. Available: <https://www.w3.org/resources/>.
- [12] CKAN, «KAN Documentation (v2.10),» [Ηλεκτρονικό]. Available: <https://docs.ckan.org/en/2.10/contents.html>. [Πρόσβαση 2024].
- [13] Open Knowledge Foundation Greece, «okfn.gr,» Open Knowledge Foundation Greece, [Ηλεκτρονικό]. Available: <https://okfn.gr/why-open/>.
- [14] Open Development Mekong, «Open Data,» 2020. [Ηλεκτρονικό]. Available: https://wiki.opendevopmentmekong.net/public:open_data.
- [15] Open Knowledge Foundation, «Open Definition 2.1,» [Ηλεκτρονικό]. Available: <https://opendefinition.org/od/2.1/en/>.
- [16] Open Knowledge Foundation, «Open Knowledge Foundation,» Open Knowledge Foundation, [Ηλεκτρονικό]. Available: <https://okfn.org/en/library/what-is-open/>.
- [17] Open Knowledge Foundation, «The Open Definition v2.1,» [Ηλεκτρονικό]. Available: <https://opendefinition.org/od/2.1/en/>.

- [18] Ευρωπαϊκή Επιτροπή, «Ευρωπαϊκή νομοθεσία για τα ανοικτά δεδομένα,» [Ηλεκτρονικό]. Available: <https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data>.
- [19] Ευρωπαϊκή Επιτροπή, «Ευρωπαϊκή στρατηγική για τα δεδομένα,» [Ηλεκτρονικό]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.
- [20] Open Knowledge Foundation Greece, «Η ΑΠΟΣΤΟΛΗ ΜΑΣ,» [Ηλεκτρονικό]. Available: <https://okfn.gr/our-mission/>.
- [21] Διαύγεια, «Πληροφορίες για το πρόγραμμα Διαύγεια,» [Ηλεκτρονικό]. Available: <https://diavgeia.gov.gr/info>.
- [22] H. Akyürek, C. Scholl, S. Regina, T. Siebenlist και A. Mainka, «Maturity and Usability of Open Data in North Rhine-Westphalia,» Düsseldorf, 2018.
- [23] S. Badiee, J. Crowell, L. Noe, A. Pittman, C. Rudow και E. Swanson, «Open Data for Official Statistics: History, Principles, and Implementation-Caleb Rudow and Eric Swanson,» Open Data Watch, Washington, DC, USA, 2021.
- [24] B. A. Lori και J. Craner, «Open Data: What It Is and Why You Should Care,» 2017. [Ηλεκτρονικό]. Available: https://www.researchgate.net/publication/317666305_Open_Data_What_It_Is_and_Why_You_Should_Care.
- [25] M. P. Page, E. Hajduk, E. N. Lincklaen Arriëns, G. Cecconi και S. Brinkhuis, «2023 Open Data Maturity Report,» <https://data.europa.eu/>, December 2023.
- [26] d. B. Patrick και P. M. PhD, «Indicators for an Open Data Impact Assessment,» European Commission Directorate-General for Communications Networks, Content and Technology, February 2024.
- [27] Α. Αθανάσιου, «Ανάπτυξη Συστήματος Σηματολογικού για την Παρακολούθηση της τήρησης του Εσωτερικού Κανονισμού του ΔΠΜΣ,» ΘΕΣΣΑΛΟΝΙΚΗ, ΙΑΝΟΥΑΡΙΟΣ 2014.
- [28] Γ. Αντωνίου και F. Van Harmelen, Εισαγωγή στο σημασιολογικό ιστό, Κλειδάριθμος, 2009.
- [29] Α. Βαβαρούτα, «Αξιοποιώντας τα ανοικτά κυβερνητικά δεδομένα για την παραγωγή,» 2022.
- [30] Π. Θωδωρής, «Ανοικτά Δημόσια Δεδομένα,» Εθνικό Κέντρο Δημόσιας Διοίκησης και Αυτοδιοίκησης.
- [31] European Data, «Linking data: Data Catalogue Vocabulary Application Profile,» European Data, 22 November 2022. [Ηλεκτρονικό]. Available: <https://data.europa.eu/en/publications/datastories/linking-data-data-catalogue-vocabulary-application-profile>. [Πρόσβαση 2024].
- [32] L. Martín, J. Lanza, V. González, J. R. Santana, P. Sotres και L. Sánchez, «A Connector for Integrating NGSI-LD Data into Open Data Portals,» MDPI, Basel, Switzerland, 2024.
- [33] ckan, «<https://docs.ckan.org/>,» ckanext-dcat, [Ηλεκτρονικό]. Available: <https://docs.ckan.org/projects/ckanext-dcat/en/latest/endpoints/>.
- [34] DKAN Open Data Portal, «dkan.readthedocs.io,» [Ηλεκτρονικό]. Available: <https://dkan.readthedocs.io/en/latest/introduction/overview.html>.
- [35] Visual Studio Code, «Visual Studio Code,» [Ηλεκτρονικό]. Available: <https://code.visualstudio.com/>.

- [36] Postman, «Postman,» [Ηλεκτρονικό]. Available: <https://www.postman.com/>.
- [37] pgadmin, «pgadmin,» [Ηλεκτρονικό]. Available: <https://www.pgadmin.org/>.
- [38] w3schools/postgresql/, «w3schools,» [Ηλεκτρονικό]. Available: <https://www.w3schools.com/postgresql/>.
- [39] Python documentation, «docs.python.org/3/,» [Ηλεκτρονικό]. Available: <https://docs.python.org/3/>.
- [40] P. Miller, R. Styles και T. Heath, «OPEN DATA COMMONS, A LICENSE FOR OPEN DATA,» Birmingham.
- [41] creativecommons.org, «creativecommons,» [Ηλεκτρονικό]. Available: <https://creativecommons.org/>.
- [42] creativecommons.ellak.gr, «creativecommons.ellak.gr,» [Ηλεκτρονικό]. Available: <https://creativecommons.ellak.gr/fylladio/>. [Πρόσβαση 2024].
- [43] Open Knowledge Foundation - OPEN DATA COMMONS, «opendatacommons.org,» Open Knowledge Foundation, [Ηλεκτρονικό]. Available: <https://opendatacommons.org/>. [Πρόσβαση 2024].
- [44] Wikipedia Creative Commons License, «Wikipedia Creative Commons License,» Wikipedia, [Ηλεκτρονικό]. Available: https://en.wikipedia.org/wiki/Creative_Commons_license. [Πρόσβαση 2024].
- [45] Open Data Commons Open Database License (ODbL), «Open Data Commons Open Database License (ODbL),» Open Knowledge Foundation, [Ηλεκτρονικό]. Available: <https://opendatacommons.org/licenses/odbl/>.
- [46] Open Data Commons Attribution License (ODC-By), «Open Data Commons Attribution License (ODC-By),» Open Knowledge Foundation, [Ηλεκτρονικό]. Available: <https://opendatacommons.org/licenses/by/>.
- [47] Open Data Commons Public Domain Dedication and License (PDDL), «Open Data Commons Public Domain Dedication and License (PDDL),» Open Knowledge Foundation, [Ηλεκτρονικό]. Available: <https://opendatacommons.org/licenses/pddl/>.
- [48] drupal.org, «drupal.org/about,» [Ηλεκτρονικό]. Available: <https://www.drupal.org/about>. [Πρόσβαση 2024].
- [49] wikipedia.org/wiki/Drupal, «wikipedia,» [Ηλεκτρονικό]. Available: <https://el.wikipedia.org/wiki/Drupal>. [Πρόσβαση 2024].
- [50] ckan.org/api/, «CKAN DOCS API GUIDE,» [Ηλεκτρονικό]. Available: <https://docs.ckan.org/en/2.8/api/>. [Πρόσβαση 2024].
- [51] wikipedia.org/wiki/API, «Wikipedia/Application Programming Interface,» [Ηλεκτρονικό]. Available: <https://en.wikipedia.org/wiki/API>.

ΠΑΡΑΡΤΗΜΑ Α : CUSTOM ΚΩΔΙΚΑΣ

ΚΩΔΙΚΑΣ PYTHON

```

1. # db_Connection.py
2. import psycopg2
3. import configparser
4.
5. def read_db_config():
6.     config_file_path = 'db.ini'
7.     config = configparser.ConfigParser()
8.     config.read(config_file_path)
9.
10.    try:
11.        ckan_db_params = config['ckan_default']
12.        datastore_db_params = config['datastore_default']
13.    except KeyError as e:
14.        print(f"Error: {e} section not found in the configuration file.")
15.        raise
16.
17.    return ckan_db_params, datastore_db_params
18.
19. def connect_to_db(db_params):
20.
21.     conn = psycopg2.connect(
22.         dbname=db_params['dbname'],
23.         user=db_params['user'],
24.         password=db_params['password'],
25.         host=db_params['host'],
26.         port=db_params['port']
27.     )
28.     return conn
29.

```

```

1. [ckan_default]
2. dbname = ckan_default
3. user = ckan_default
4. password = *
5. host = localhost
6. port = 5432
7.
8. [datastore_default]
9. dbname = datastore_default
10. user = ckan_default
11. password = *
12. host = localhost
13. port = 5432
14.

```

```

1.
2. # WaterQuality_To_Datastore.py
3. import os
4. import csv
5. import json
6. import requests
7. import psycopg2
8. import configparser
9. import db_Connection
10. from bs4 import BeautifulSoup
11.
12.

```

```

13. # Ζητά τον σύνδεσμο από τον χρήστη
14. main_url = "https://quality.eyath.gr/" #input("Εισάγετε Link (URL): ")
15.
16. class WaterQualityScrape:
17.     def __init__(self, url):
18.         self.url = url
19.         # self.processed_links = [] # Λίστα για την αποθήκευση των συνδέσμων που έχουν
επεξεργαστεί
20.
21.     def scrape_categories(self):
22.
23.         # Στέλνει GET αίτημα στο URL στο self.url
24.         response = requests.get(self.url)
25.
26.         # Έλεγχος για επιτυχής απάντηση
27.         if response.status_code != 200:
28.             print({response.status_code})
29.             return []
30.
31.         # Χρησιμοποιεί τη βιβλιοθήκη BeautifulSoup για HTML περιεχόμενο της απάντησης
32.         soup = BeautifulSoup(response.text, 'html.parser')
33.
34.         # Βρίσκει όλα τα 'div' στοιχεία που έχουν την κλάση 'home-cats'
35.         # <div class="elementor-shortcode"><div class="home-cats col-md-4">
36.         # <a href="URL">40 Εκκλησιές</a></div>
37.         category_links = soup.find_all('div', class_='home-cats')
38.
39.         #Για Αποθήκευση Περιοχών Link,ονομα
40.         categories = []
41.
42.         # Επαναλαμβάνει για κάθε 'div' στοιχείο που βρέθηκε με την κλάση 'home-cats'
43.         for link in category_links:
44.
45.             # Αποθηκεύει το URL της κατηγορίας που βρίσκεται στο href του 'a' στοιχείου μέσα
στο 'div'
46.             url_category = link.a['href']
47.
48.             # Παίρνει το ονομα της περιοχής
49.             name_category = link.a.get_text().strip()
50.             # Εξαιρεί την κατηγορία που περιέχει "Σύνολο Αναλύσεων Δικτύου Πόλης"
51.             if "Σύνολο Αναλύσεων Δικτύου Πόλης" in name_category:
52.                 continue
53.
54.             #Αποθήκευση link , category περιοχής στο categories = []
55.             categories.append({"url": url_category, "name": name_category})
56.
57.         return categories
58.
59.     def scrape_month_year(self, url):
60.         # Στέλνει GET αίτημα στο URL στο self.url
61.         response = requests.get(url)
62.
63.         # Έλεγχος για επιτυχής απάντηση
64.         if response.status_code != 200:
65.             print( {response.status_code})
66.             return []
67.
68.         # Χρησιμοποιεί τη βιβλιοθήκη BeautifulSoup για HTML περιεχόμενο της απάντησης
69.         soup = BeautifulSoup(response.text, 'html.parser')
70.
71.         # Βρίσκει όλα τα 'article' στη σελίδα
72.         # <article class="elementor-post elementor-grid-item post-912 post type-post status-
publish format-standard hentry category-40">
73.         # <div class="elementor-post__text">
74.         # <h6 class="elementor-post__title">
75.         # <a href="https://quality.eyath.gr/2024_04_saranta_ekklisies/" >
76.         # 2024 Απρίλιος </a>
77.         articles = soup.find_all('article')
78.
79.         #Για αποθήκευση πληροφορίας

```

```

80.         results = []
81.
82.         # για κάθε 'article' στοιχείο που βρέθηκε στη σελίδα
83.         for article in articles:
84.             # Αποθηκεύει το URL που βρίσκεται στο href του 'a' στοιχείου μέσα στο 'article'
85.             link = article.find('a')['href']
86.             # Παίρνει το κείμενο <a> καθαρίζει κενά και χαρακτήρες και το αποθηκεύει
87.             month_year = article.find('a').text.strip().replace('\xa0', ' ')
88.
89.             # Προσθέτει το URL μήνα και έτος στον results
90.             results.append({"Link": link, "Month and Year": month_year})
91.
92.         return results
93.
94.     def scrape_table(self, url):
95.         response = requests.get(url)
96.         if response.status_code != 200:
97.             print(f"Η αίτηση για το URL '{url}' δεν ήταν επιτυχής. Κωδικός κατάστασης: {response.status_code}")
98.             return []
99.
100.        soup = BeautifulSoup(response.text, 'html.parser')
101.        tables = soup.find_all('table')
102.
103.        all_data = []
104.
105.        for table in tables:
106.            headers = [header.get_text().strip() for header in table.find_all('th')]
107.
108.            # Καθορισμός του τύπου παραμέτρου από την πρώτη λέξη της πρώτης επικεφαλίδας
109.            TypeHeader = headers[0]
110.            parameter_type = TypeHeader.split()[0] # Παίρνει την πρώτη λέξη της πρώτης
επικεφαλίδας
111.
112.            # Βρίσκουμε τη θέση της στήλης "Τιμή"
113.            try:
114.                Header_timi = headers.index("Τιμή")
115.            except ValueError:
116.                print(f"Η στήλη 'Τιμή' δεν βρέθηκε στον πίνακα με επικεφαλίδες: {headers}")
117.                continue
118.
119.            for row in table.find_all('tr')[1:]: # Skip header row
120.                columns = row.find_all('td')
121.
122.                parameter_name = columns[0].get_text().strip()
123.
124.                for sup_tag in columns[Header_timi].find_all('sup'):
125.                    sup_tag.decompose()
126.
127.                parameter_value_timi = columns[Header_timi].get_text().strip()
128.                ##parameter_value_timi = columns[Header_timi].get_text(separator='^').strip()
129.                if columns[Header_timi].find('sup') else columns[Header_timi].get_text().strip()
130.                if parameter_value_timi == "":
131.                    parameter_value_timi = "null"
132.
133.                all_data.append((parameter_type, parameter_name, parameter_value_timi))
134.
135.        return all_data

```

```

1. class CkanRequest:
2.
3.     ##Εδώ Περνουμε ολα τα δεδομενα για να τα κανουμε insert σε temp Πινακα για την συγκριση
μετα
4.     def Get_data_all(api_key, resource_id):
5.         get_url = 'http://127.0.0.1:5000/api/3/action/datastore_search'
6.         headers = {'X-CKAN-API-Key': api_key}
7.
8.         records = []

```

```

9.     offset = 0
10.    limit = 100000000000 # limit
11.
12.    while True:
13.        params = {
14.            'resource_id': resource_id,
15.            'offset': offset,
16.            'limit': limit
17.        }
18.
19.        response = requests.get(get_url, headers=headers, params=params)
20.
21.        if response.status_code == 200:
22.            data = response.json()
23.            records = data.get('result', {}).get('records', [])
24.            records.extend(records)
25.
26.            # Ελεγχει για εξτρα δεδομενα
27.            if len(records) < limit:
28.                break
29.            else:
30.                offset += limit
31.
32.        else:
33.            print(f"Σφάλμα Ανακτησεις Δεδομένων απο το CKAN. Status code:
34.            {response.status_code}")
35.            break
36.
37.    return records
38.
39.    def insert_new_data(api_key, resource_id, new_data):
40.        datastore_data = {
41.            'resource_id': resource_id,
42.            'records': new_data,
43.            'method': 'insert',
44.            "force": "True"
45.        }
46.
47.        try:
48.            response = requests.post(
49.                f'http://127.0.0.1:5000/api/3/action/datastore_upsert',
50.                headers={'X-CKAN-API-Key': api_key},
51.                json=datastore_data
52.            )
53.
54.            response.raise_for_status()
55.
56.            if response.status_code == 200:
57.                print('Υπάρχουν Νέα Δεδομένα.')
58.            else:
59.                print('ΣΦΑΛΜΑ Inserted/Updates data in CKAN:', response.text)
60.
61.        except requests.exceptions.RequestException as e:
62.            print(f"ΣΦΑΛΜΑ Inserted/Updates data in CKAN: ' {e}'")
63.
64.    def create_datastore(api_key, fields, records):
65.
66.        year_month_records = {}
67.
68.        for record in records:
69.            year = str(record["Year"]).replace(" ", "_")
70.            month = record["Month"].replace(" ", "_")
71.            category = record["Category"].replace(" ", "_")
72.            type_ = record["Type"].replace(" ", "_")
73.            key = f"{year}_{month}_{category}_{type_}"
74.
75.            if key not in year_month_records:
76.                year_month_records[key] = []
77.

```

```

78.         year_month_records[key].append(record)
79.
80.     create_url = 'http://127.0.0.1:5000/api/3/action/datastore_create'
81.     headers = {'Content-Type': 'application/json', 'X-CKAN-API-Key': api_key}
82.
83.     for key, records in year_month_records.items():
84.         resource = {
85.             "resource": {
86.                 "package_id": 'e24f0295-06aa-410a-8087-e6aaa5011a70',
87.                 "name": key,
88.                 "datastore_name": key,
89.                 "datastore_active": True,
90.                 "format": "CSV"
91.             },
92.             "fields": fields,
93.             "records": records,
94.             "method": "insert",
95.             "force": True
96.         }
97.
98.
99.         print(f"Request για {key} με Δεδομένα:")
100.        print(json.dumps(resource, indent=2))
101.
102.        response = requests.post(create_url, headers=headers, data=json.dumps(resource))
103.
104.        if response.status_code == 200:
105.            print(f"To datastore δημιουργήθηκε για {key}")
106.        else:
107.            print(f"Ανεπιτυχής δημιουργία datastore για {key}. Status code:
108.            {response.status_code}")
109.            print(response.text)

```

```

1. class ExecuteSql:
2.     def ExeSql():
3.
4.
5.         ckan_db_params, datastore_db_params = db_Connection.read_db_config()
6.
7.
8.         ckan_conn = db_Connection.connect_to_db(ckan_db_params)
9.         datastore_conn = db_Connection.connect_to_db(datastore_db_params)
10.
11.
12.         ckan_cur = ckan_conn.cursor()
13.         datastore_cur = datastore_conn.cursor()
14.
15.
16.         ## Εδώ αυτή η Function Απλά κάνει insert στον πίνακα resource_view για να εμφανίζετε
17.         η View στο Ckan Τα νέα Datastores
18.         ckan_cur.execute('SELECT public.al_insertViews()')
19.         ckan_conn.commit()
20.         print("Function al_insertViews() Εκτελέστηκε στην Βάση του CKAN.")
21.
22.         ckan_cur.execute('CALL public.al_updateTimeToMainDatastore()')
23.         ckan_conn.commit()
24.         print("Procedure al_updateTimeToMainDatastore() Εκτελέστηκε στην Βάση του DATASTORE.")
25.
26.
27.         ## Η procedure Αυτή Κάνει Insert Τα νέα Δεδομένα που έχουμε πχ Ιουνία 2024 θα βρει για
28.         Τύπο και Περιοχή Water Quality Group By Area (Thessaloniki) & Type
29.         # και θα κάνει insert τον καινούργιο Μήνα στους πίνακες όπου είναι τα δεδομένα
30.         αυτά
31.         datastore_cur.execute('CALL public.al_procedureinsertttoresources()')
32.         datastore_conn.commit()
33.         print("Procedure al_procedureinsertttoresources() Εκτελέστηκε στην Βάση του CKAN.")

```

```

33.         ckan_cur.close()
34.         ckan_conn.close()
35.         datastore_cur.close()
36.         datastore_conn.close()
37.

```

```

1. if __name__ == "__main__":
2.     #To api_key στο CKAN.
3.
4.     # To resource_id αφορά τον συγκεκριμένο πίνακα μέσα στο CKAN DataStore
5.     resource_id = "9dd8393f-9728-40a1-96f6-168241e953ae"
6.     #Ορισμός των πεδίων (fields) που πιθανόν χρειάζονται για τη δημιουργία του DataStore
7.     fields = [
8.         {"id": "Year", "type": "text"},
9.         {"id": "Month", "type": "text"},
10.        {"id": "Category", "type": "text"},
11.        {"id": "Type", "type": "text"},
12.        {"id": "Parameter", "type": "text"},
13.        {"id": "Value", "type": "text"}
14.    ]
15.
16.    try:
17.        # Καλούμε τη μέθοδο Get_data_all() του CkanRequest για να πάρουμε όλα τα αρχεία
(records)
18.        # που υπάρχουν στο resource με ID = resource_id.
19.        records = CkanRequest.Get_data_all(api_key, resource_id)
20.
21.        #Μετατρέπουμε τη λίστα των records σε ένα Python set από tuples,
22.        # έτσι ώστε να μπορούμε εύκολα να ελέγχουμε αν κάποιο record υπάρχει ήδη.
23.        # Παράδειγμα tuple = ('2023', 'July', 'CategoryName', 'TypeName', 'ParameterName',
'123')
24.        existing_records_set = {(
25.            str(rec['Year']).strip(),
26.            str(rec['Month']).strip(),
27.            rec['Category'].strip(),
28.            rec['Type'].strip(),
29.            rec['Parameter'].strip(),
30.            str(rec['Value']).strip())
31.            for rec in records}
32.
33.        # Εκτύπωση όλων των υπάρχοντων records για δικούς μου ελεγχους
34.        print("Υπάρχοντα records:")
35.        for rec in existing_records_set:
36.            print(rec)
37.        # Εκτύπωση του συνολικού αριθμού των record που υπάρχουν ήδη για δικούς μου
ελεγχους
38.        print(f"\nΣυνολικά Υπαρχοντα records: {len(existing_records_set)}")
39.
40.        #Δημιουργούμε ένα αντικείμενο scraper (WaterQualityScrape),
41.        # το οποίο έχει την μέθοδο scrape_categories() στην κλάση WaterQualityScrape για να
πάρει κατηγορίες/περιοχές.
42.        scraper = WaterQualityScrape(main_url)
43.        categories = scraper.scrape_categories()
44.
45.        # Εδώ θα αποθηκεύσουμε τα νέα δεδομένα (records)
46.        new_data_to_insert = []
47.
48.
49.        # Για κάθε κατηγορία που βρέθηκε στο scraping
50.        for category in categories:
51.            # Εκτυπώνουμε το όνομα της περιοχής/κατηγορία
52.            print(f"\nΠεριοχή: {category['name']}")
53.            # Καλούμε τη μέθοδο scrape_month_year() για να πάρουμε τις πληροφορίες
54.            # σχετικά με τον μήνα και το έτος, από το URL της κατηγορίας
55.            data = scraper.scrape_month_year(category['url'])
56.
57.            # Για κάθε στοιχείο (item) που περιέχει πληροφορίες για Month και Year

```

```

58.         for item in data:
59.             # Κόβουμε το string "Month Year" σε δύο μέρη (π.χ. "2023 June")
60.             parts = item['Month and Year'].split()
61.             year, month = parts
62.             # Παίρνουμε τα αναλυτικά δεδομένα (tables_data) από scrape_table
63.             tables_data = scraper.scrape_table(item['Link'])
64.             print(f"\nΠεριοχή: {category['name']} Έτος: {year} Μήνας: {month}")
65.
66.             # 5. Έλεγχος αν τα νέα δεδομένα υπάρχουν ήδη στο set
67.             for type_, parameter, value in tables_data:
68.                 # Φτιάχνουμε ένα tuple (record_key) που είναι ο συνδυασμός των πεδίων
69.                 record_key = (year.strip(), month.strip(), category['name'].strip(),
type_.strip(), parameter.strip(), value.strip())
70.                 # Αν το record_key ΔΕΝ υπάρχει ήδη στα existing_records_set...
71.                 if record_key not in existing_records_set:
72.                     # φτιάχνουμε ένα καινούργιο record σε JSON μορφή
73.                     new_record = {
74.                         "Year": year.strip(),
75.                         "Month": month.strip(),
76.                         "Category": category['name'].strip(),
77.                         "Type": type_.strip(),
78.                         "Parameter": parameter.strip(),
79.                         "Value": value.strip()
80.                     }
81.                     # Το προσθέτουμε στη λίστα new_data_to_insert
82.                     new_data_to_insert.append(new_record)
83.                     # Και ενημερώνουμε το existing_records_set για να μην το ξαναβρούμε
διπλό
84.                     existing_records_set.add(record_key)
85.
86.
87.             print("Νέα Δεδομένα για εισαγωγή")
88.             for rec in new_data_to_insert:
89.                 print(rec)
90.
91.             if new_data_to_insert:
92.
93.                 CkanRequest.insert_new_data(api_key, resource_id, new_data_to_insert)
94.
95.                 CkanRequest.create_datastore(api_key, fields, new_data_to_insert)
96.                 print(f"Έγινε Είσαγωγή {len(new_data_to_insert)} Νέων Δεδομένων.")
97.             else:
98.                 print("Δεν υπάρχουν καινούργια δεδομένα για εισαγωγή")
99.
100.            print("Είσαγωγή ολοκληρώθηκε.")
101.
102.            ##Εκτέλεση SQL για Ενημέρωση των Datastore Water Quality Group By Area (Thessaloniki) &
Type
103.            ExecuteSql.ExeSql()
104.
105.        except Exception as e:
106.            print(f"Error: {str(e)}")
107.
108.
109.

```

View.py – Custom Extension Ckanext-WaterThessaloniki

```

1. import subprocess
2. import os
3. import signal
4. from threading import Lock
5. import threading
6. from flask import Blueprint, jsonify
7. import logging
8.
9. # Δημιουργία ενός Flask Blueprint με όνομα "waterthessaloniki" και URL prefix "/waterthessaloniki"
10. waterthessaloniki = Blueprint("waterthessaloniki", __name__, url_prefix="/waterthessaloniki")
11.

```

```

12. # Αποθήκευση της κατάστασης του script (αν τρέχει ή όχι) και άλλες σχετικές πληροφορίες
13. status_info = {"pid": None, "status": "idle", "output": ""} # idle, running
14. lock = Lock()
15.
16. script_path = "/home/aggelospincha/ckan/lib/default/src/ckan/ckan/CustomScripts/WaterQuality_To_DataStore.py" # Update
with actual path
17.
18. @waterthessaloniki.route("/start", methods=["POST"])
19. def start_script():
20.     global status_info
21.     with lock: # Κλείδωμα για αποφυγή ταυτόχρονων τροποποιήσεων
22.         if status_info["status"] == "running": # Έλεγχος αν το script είναι ήδη σε λειτουργία
23.             return jsonify({"message": "Script is already running."}), 400
24.         # Εντολή για ενεργοποίηση του CKAN virtual environment και εκτέλεση του script
25.         command = f". /usr/lib/ckan/default/bin/activate && python3 {script_path}"
26.         process = subprocess.Popen(
27.             command, # Εντολή που θα εκτελεστεί
28.             stdout=subprocess.PIPE, # Αποθήκευση εξόδου (stdout)
29.             stderr=subprocess.PIPE, # Αποθήκευση σφαλμάτων (stderr)
30.             shell=True, # Εκτέλεση εντολής μέσω shell
31.             preexec_fn=os.setsid # Δημιουργία νέας ομάδας διεργασιών για ευκολότερο τερματισμό
32.         )
33.
34.         # Αποθήκευση του PID και ενημέρωση της κατάστασης
35.         status_info["pid"] = process.pid
36.         status_info["status"] = "running"
37.
38.         # Συνάρτηση για παρακολούθηση του script (μέσω thread)
39.         def wait_for_process():
40.             stdout, stderr = process.communicate() # Περιμένει την ολοκλήρωση του script και λαμβάνει την έξοδο
41.             output = stdout.decode() + stderr.decode() # Συνδυασμός εξόδου και σφαλμάτων σε string
42.
43.             # Αποθήκευση εξόδου και επιστροφή κατάστασης σε "idle"
44.             status_info["output"] = output
45.             status_info["status"] = "idle" # Update status to idle when done
46.
47.             # Log all the output and error in the log file
48.             logging.info(f"Script output:\n{stdout.decode()}")
49.
50.             status_info["status"] = "idle" # Update status to idle when done
51.
52.
53.             logging.info(f"Script completed successfully with PID {status_info['pid']}".)
54.
55.         # Δημιουργία thread για παρακολούθηση του script
56.         threading.Thread(target=wait_for_process).start()
57.
58.         # Επιστροφή απάντησης με το PID και την έξοδο
59.         return jsonify({
60.             "message": "Script started.",
61.             "pid": status_info["pid"],
62.             "output": status_info["output"]
63.         })
64.
65. # Route για τον τερματισμό του script
66. @waterthessaloniki.route("/terminate", methods=["POST"])
67. def terminate_script():
68.     global status_info
69.     with lock:
70.         if status_info["status"] != "running": # Κλείδωμα για αποφυγή ταυτόχρονων τροποποιήσεων
71.             return jsonify({"message": "No running script to terminate."}), 400
72.
73.         try:
74.             # Τερματισμός διεργασίας μέσω PID μέσω το killpg παίρνώντας το pid από το Process_info
75.             os.killpg(os.getpgid(status_info["pid"]), signal.SIGTERM)
76.             status_info["status"] = "idle" # Ενημέρωση κατάστασης
77.             status_info["pid"] = None # Καθαρισμός PID
78.
79.             return jsonify({"message": "Script terminated."})
80.
81.         except ProcessLookupError:

```

```

82.     status_info["status"] = "idle"
83.     status_info["pid"] = None
84.     return jsonify({"message": "Process not found."}), 404
85.
86. # Route για έλεγχο της κατάστασης του script
87. @waterthessaloniki.route("/status", methods=["GET"])
88. def get_status():
89.     return jsonify(status_info) # Επιστροφή πληροφοριών για την τρέχουσα κατάσταση
90.
91. # Συνάρτηση για επιστροφή των blueprints του CKAN extension
92. def get_blueprints():
93.     """
94.     Return the blueprints for the CKAN extension.
95.     """
96.     return [waterthessaloniki]
97.

```

SQL ΚΩΔΙΚΑΣ

```

1. -- FUNCTION: public.al_insertViews()
2.
3. -- DROP FUNCTION IF EXISTS public."al_insertViews"();
4.
5. CREATE OR REPLACE FUNCTION public."al_insertViews"(
6. )
7. RETURNS void
8. LANGUAGE 'plpgsql'
9. COST 100
10. VOLATILE PARALLEL UNSAFE
11. AS $BODY$
12. DECLARE
13.
14.     -- 1) resource_rec: RECORD, θα κρατάει προσωρινά τα δεδομένα από το SELECT
15.     -- 2) view_id: UUID, για να αποθηκεύουμε το id του νέου view που θα δημιουργείται
16.     resource_rec RECORD;
17.     view_id UUID;
18. BEGIN
19.     -- Κανουμε select (r.id, r.name) από τον πίνακα resource (r)
20.     -- μέσω LEFT JOIN με το resource_view (rv).
21.     -- Η συνθήκη rv.id IS NULL σημαίνει ότι δεν βρέθηκε καμία εγγραφή στον πίνακα resource_view
22.     -- για τον συγκεκριμένο resource_id (άρα δεν υπάρχουν views).
23.     -- Το r.state = 'active' σημαίνει ότι φιλτράρουμε μόνο τους ενεργούς πόρους.
24.     FOR resource_rec IN
25.         SELECT r.id AS resource_id, r.name AS title
26.         FROM resource r
27.         LEFT JOIN resource_view rv ON r.id = rv.resource_id
28.         WHERE rv.id IS NULL AND r.state = 'active'
29.     LOOP
30.         --Δημιουργούμε ένα νέο UUID για το resource_view που θα εισαχθεί
31.         view_id := uuid_generate_v4();
32.
33.         -- Κάνουμε INSERT στον πίνακα resource_view,
34.         -- βάζοντας τα βασικά πεδία ώστε να δημιουργηθεί
35.         INSERT INTO resource_view (id, resource_id, title, description, view_type, "order",
36.         config)
37.         VALUES (view_id, resource_rec.resource_id, resource_rec.title, 'Automatically created
38.         view', 'recline_view', 1, null);
39.         -- δείχνοντας το id του νέου view και τον αντίστοιχο resource_id.
40.         RAISE NOTICE 'Inserted resource view with id % for resource_id: %', view_id,
41.         resource_rec.resource_id;
42.     END LOOP;
43.     RAISE NOTICE 'Completed inserting missing resource views.';
44. END;
45. $BODY$;

```

```

45.
46. ALTER FUNCTION public."al_insertViews"()
47.     OWNER TO ckan_default;
48.
49.

```

```

1. PROCEDURE al_updateTimeToMainDatastore
2.
3. CREATE OR REPLACE PROCEDURE public.al_updatetimetomaindatastore(
4.     )
5. LANGUAGE 'plpgsql'
6. AS $BODY$
7. BEGIN
8.     UPDATE public.resource
9.     SET metadata_modified = CURRENT_TIMESTAMP,
10.        last_modified = CURRENT_TIMESTAMP
11.     WHERE id = '9dd8393f-9728-40a1-96f6-168241e953ae';
12. END;
13. $BODY$;
14. ALTER PROCEDURE public.al_updatetimetomaindatastore()
15.     OWNER TO ckan_default;
16.
17. FUNCTION public.al_insert_combined_results -DATASTORE
18.

```

```

1. CREATE OR REPLACE FUNCTION public.al_insert_combined_results(
2.     )
3. RETURNS void
4. LANGUAGE 'plpgsql'
5. COST 100
6. VOLATILE PARALLEL UNSAFE
7. AS $BODY$
8. DECLARE
9.     current_resourceid TEXT;
10.    sql TEXT;
11. BEGIN
12.
13.     DROP TABLE IF EXISTS combined_results;
14.
15.
16.     CREATE TEMP TABLE combined_results (
17.         _id INT,
18.         _full_text TEXT,
19.         "Year" INTEGER,
20.         "Month" TEXT,
21.         "Category" TEXT,
22.         "Type" TEXT,
23.         "Parameter" TEXT,
24.         "Value" TEXT,
25.         resourceid UUID
26.     );
27.
28.     --Εύρεση (distinct) resource IDs από τον πίνακα "9dd8393f-9728-40a1-96f6-168241e953ae"
29.     -- μέσω LEFT JOIN με τη συνάρτηση dblink (al_ResourcesCategoryType).
30.     --
31.     -- - To dblink καλείται με 'SELECT resourceid, category, "type" FROM
al_ResourcesCategoryType',
32.     -- επιστρέφοντας στήλες (resourceid TEXT, category TEXT, "type" TEXT).
33.     -- - Γίνεται JOIN με local_table."Category" = al_resourceID.category
34.     -- και local_table."Type" = al_resourceID."type".
35.     -- - Στα σχόλια υπάρχουν γραμμές WHERE που μπορείτε να ενεργοποιήσετε για
36.     -- να φιλτράρετε συγκεκριμένο Έτος, Μήνα, Category κ.λπ.
37.     FOR current_resourceid IN
38.
39.         SELECT DISTINCT resourceid
40.         FROM (
41.             SELECT DISTINCT al_resourceID.resourceid
42.             FROM public."9dd8393f-9728-40a1-96f6-168241e953ae" AS local_table

```

```

43.         LEFT JOIN dblink(
44.             'dbname=ckan_default',
45.             'SELECT resourceid, category, "type" FROM al_ResourcesCategoryType'
46.         ) AS al_resourceID(resourceid TEXT, category TEXT, "type" TEXT)
47.         ON al_resourceID.category = local_table."Category"
48.         AND al_resourceID."type" = local_table."Type"
49.         -- WHERE local_table."Year" = 2024
50.         -- AND local_table."Month" = 'Ιούνιος'
51.         -- AND local_table."Category" = '40 Εκκλησιές'
52.     ) AS resource_ids
53. LOOP
54.     -- `` Για κάθε resourceid που βρέθηκε, φτιάχνουμε δυναμικό SQL που:
55.     -- - Κάνει INSERT INTO TEMP combined_results
56.     -- - Επιλέγει (SELECT) τις στήλες που θέλουμε από τον πίνακα
57.     -- - Χρησιμοποιούμε quote_literal(current_resourceid) ως resourceid
58.     sql := 'INSERT INTO combined_results ' ||
59.         'SELECT _id, _full_text, CAST("Year" AS INTEGER), "Month", "Category", "Type",
60.         "Parameter", "Value", ' ||
61.         quote_literal(current_resourceid) || ' AS resourceid ' ||
62.         'FROM public."' || current_resourceid || ''';
63.
64.     RAISE NOTICE '%', sql;
65.
66.     EXECUTE sql;
67. END LOOP;
68.
69.
70.
71. RAISE NOTICE 'Combined Results: ';
72.
73. PERFORM * FROM combined_results;
74. END;
75. $BODY$;
76.

```

```

1. CREATE OR REPLACE FUNCTION public.al_insert_into_rid_tables(
2.     )
3.
4. -- Περιγραφή:
5. -- Η συνάρτηση insert_into_rid_tables():
6. -- Διατρέχει (με FOR LOOP) μια λίστα από εγγραφές (current_row) που προκύπτουν
7. -- από ένα SELECT σε πίνακες "9dd8393f-9728-40a1-96f6-168241e953ae", combined_results
8. -- και ένα υπο-SELECT (v).
9. -- Για κάθε τέτοια εγγραφή, κάνει INSERT
10. -- που εισάγει τα σχετικά δεδομένα στον πίνακα public."rid" resources_ID
11. --
12. -- Μετά το INSERT, καλεί μια dblink_exec για να ενημερώσει τον πίνακα resource
13. -- αλλάζοντας timestamps (metadata_modified κ.λπ.).
14. -- Ουσιαστικά, εισαγει νέες εγγραφές σε διαφορετικούς πίνακες με βάση
15. -- το πεδίο rid (το rid είναι και το όνομα του πίνακα στον πίνακα του datastore).
16. --
17.
18. -- - Ο πίνακας "9dd8393f-9728-40a1-96f6-168241e953ae" είναι ο κύριο resource πίνακας.
19. -- - Ο πίνακας combined_results είναι προσωρινός ή μόνιμος πίνακας
20. -- που έχουμε από τη συνάρτηση al_insert_combined_results().
21.
22.
23.
24. RETURNS void
25. LANGUAGE 'plpgsql'
26. COST 100
27. VOLATILE PARALLEL UNSAFE
28. AS $BODY$
29. DECLARE
30.     current_row RECORD;
31.     sql TEXT;

```

```

32. BEGIN
33.   --Παιρνουμε τα δεδομένα απο τον κυριο πινακα που δλδ--Νεος μηνας δεν ειναι μεσα στον πινακα
του εκαστοτε
34.   --Πινακα των View παρνουμε τα δεδομένα που μας λειπουν και τα κανουμε insesrt με την
βοηθεια του Combined_table απο την
35.   --Fuction al_insert_combined_results()
36.   FOR current_row IN
37.     SELECT DISTINCT a._id,
38.                      a._full_text,
39.                      a."Year",
40.                      a."Month",
41.                      a."Category",
42.                      a."Type",
43.                      a."Parameter",
44.                      a."Value",
45.                      v.rid
46.     FROM public."9dd8393f-9728-40a1-96f6-168241e953ae" a
47.     LEFT JOIN combined_results b
48.     ON a."Year" = b."Year"
49.        AND a."Month" = b."Month"
50.        AND a."Category" = b."Category"
51.        AND a."Type" = b."Type"
52.        AND a."Parameter" = b."Parameter"
53.     INNER JOIN (
54.       SELECT DISTINCT a."Category" AS category,
55.                        b.resourceid AS rid,
56.                        a."Type" AS type
57.       FROM public."9dd8393f-9728-40a1-96f6-168241e953ae" a --Κύριο Datastore με ολα τα
δεομένα
58.       LEFT JOIN combined_results b
59.       ON a."Year" = b."Year"
60.          AND a."Month" = b."Month"
61.          AND a."Category" = b."Category"
62.          AND a."Type" = b."Type"
63.          AND a."Parameter" = b."Parameter"
64.       WHERE b.resourceid IS NOT NULL
65.     ) v
66.     ON v.category = a."Category"
67.        AND v.type = a."Type"
68.     WHERE b._id IS NULL
69.   LOOP
70.     --Insert για καθε row_Rid
71.     sql := 'INSERT INTO public."' || current_row.rid || '" ' ||
72.            '(_full_text, "Year", "Month", "Category", "Type", "Parameter", "Value") ' ||
73.            'VALUES (' ||
74.            quote_literal(current_row._full_text) || ', ' ||
75.            current_row."Year" || ', ' ||
76.            quote_literal(current_row."Month") || ', ' ||
77.            quote_literal(current_row."Category") || ', ' ||
78.            quote_literal(current_row."Type") || ', ' ||
79.            quote_literal(current_row."Parameter") || ', ' ||
80.            quote_literal(current_row."Value") || ')';
81.
82.
83.
84.     EXECUTE sql;
85.     PERFORM dblink_exec(
86.       'dbname=ckan_default',
87.       'UPDATE resource
88.        SET
89.          metadata_modified = CURRENT_TIMESTAMP,
90.          last_modified = CURRENT_TIMESTAMP
91.        WHERE id = ''' || current_row.rid || ''' AND state = ''active'''
92.     );
93.   END LOOP;
94. END;
95. $BODY$;
96.

```

```

1. PROCEDURE al_procedureinserttoresources()
2.
3.
4. CREATE OR REPLACE PROCEDURE public.al_procedureinserttoresources(
5. )
6. LANGUAGE 'plpgsql'
7. AS $BODY$
8. BEGIN
9.     -- Εκτέλεση fucnction al_insert_combined_results
10.    PERFORM al_insert_combined_results();
11.
12.    -- Εκτέλεση function al_insert_into_rid_tables
13.    PERFORM al_insert_into_rid_tables();
14.
15.    -- Διαγραφή Πίνκαα
16.    DELETE FROM combined_results;
17.
18.
19.    RAISE NOTICE 'Procedure executed successfully.';
20. END;
21. $BODY$;
22. ALTER PROCEDURE public.al_procedureinserttoresources()
23.    OWNER TO ckan_default;
24.
25.

```

ΚΩΔΙΚΑΣ JAVASCRIPT-HTML (Front-End)

```

1.  {% block package_resources %}
2.  {% set page_size = 10 %} {% Επιλέγουμε μέγεθος σελίδας #}
3.  {% set total_resources = pkg.resources | length %}
4.
5.  <!-- Αναζήτηση -->
6.  <div style="margin-bottom: 10px;">
7.      <input type="text"
8.          id="resourceSearch"
9.          placeholder="Αναζήτηση (π.χ. 2024; Νοέμβριος ή 2024 40 Εκκλησίες)"
10.         style="width: 100%; padding: 8px;">
11.  </div>
12.
13.  <!-- CSV APXEIA -->
14.  <div id="resource-container">
15.      {% snippet "package/snippets/resources_list.html", pkg=pkg, resources=pkg.resources %}
16.  </div>
17.
18.  <!-- Σελιδοδερικτης -->
19.  <div id="pagination-controls"
20.      class="pagination-container text-center mt-3">
21.
22.      <button id="prevPage"
23.          class="btn btn-secondary"
24.          disabled>Προηγούμενο</button>
25.
26.      <span id="pageNumber">Σελίδα 1 Από 1</span>
27.
28.      <button id="nextPage" class="btn btn-primary" disabled>Επόμενο</button>
29.  </div>
30.
31.  <!-- Μεταβαση στην Σελίδα -->
32.  <div id="jump-controls" class="text-center mt-3">
33.      <div class="form-inline d-flex justify-content-center align-items-center">
34.          <label for="jumpPageInput" class="control-label mr-2">Σελίδα</label>
35.          <input
36.              type="number"

```

```

37.     id="jumpPageInput"
38.     class="form-control mr-2"
39.     min="1"
40.     step="1"
41.     placeholder=" "
42.     style="width: 50px; text-align: center;"
43.   />
44.   <button id="jumpPageButton" class="btn btn-primary">Go</button>
45. </div>
46. </div>
47.
48. <script> //εχει πλήρως φορτωθεί και δημιουργηθεί το αντίστοιχο DOM
49.   document.addEventListener('DOMContentLoaded', function () {
50.     const pageSize = parseInt("{ { page_size|int }}"); //Παίρνει την Τιμή απο το {% set
page_size = 10 %} {# Επιλέγουμε μέγεθος σελίδας #}
51.     let allItems = Array.from(document.querySelectorAll('.resource-item')); //Απο την κλάση
`.resource-item` από το DOM τα αποθηκεύει σε έναν πίνακα.
52.     let filteredItems = allItems.slice(); // Αρχικά, όλα τα στοιχεία περιλαμβάνονται
53.     let currentPage = 1; //CurentPage Μεταβλητη για τον υπολογισμο των Σελιδων
54.
55.     // Εμφάνιση δεδομένων για την τρέχουσα σελίδα Funcion RenderPage
56.     function renderPage() {
57.       allItems.forEach(item => item.style.display = 'none'); // Κρύβουμε όλα τα στοιχεία
58.
59.       const startIndex = (currentPage - 1) * pageSize; // Ξεκινάμε από το αρχικό στοιχείο
Resource Arceio kai tha emfanisoume alla 10 (Xls).
60.       const endIndex = startIndex + pageSize; //to end Index Πάιρνει τον το index απο
πάνω και απο το pagesize
61.       const pageItems = filteredItems.slice(startIndex, endIndex); //Εξάγει τα στοιχεία
για την τρέχουσα σελίδα από τον "filteredItems".
62.
63.       pageItems.forEach(item => item.style.display = ''); // Εμφανίζουμε μόνο τα στοιχεία
της σελίδας
64.
65.       //Υπολογίζει τον συνολικό αριθμό σελίδων με βάση τα φιλτραρισμένα στοιχεία και το
page Size.
66.       const totalPages = Math.ceil(filteredItems.length / pageSize);
67.
68.       //Απενεργοποιεί το κουμπί "Προηγούμενο" εάν ο χρήστης βρίσκεται στην πρώτη σελίδα.
69.       document.getElementById('prevPage').disabled = (currentPage === 1);
70.
71.       //Ενημερώνει την εμφάνιση του αριθμού σελίδας με την τρέχουσα σελίδα και το σύνολο
των σελίδων.
72.       document.getElementById('pageNumber').textContent = `Page ${currentPage} of
${totalPages}`;
73.
74.       //Απενεργοποιεί το κουμπί "Επόμενο" εάν ο χρήστης βρίσκεται στην τελευταία σελίδα
ή αν δεν υπάρχουν σελίδες.
75.       document.getElementById('nextPage').disabled = (currentPage === totalPages ||
totalPages === 0);
76.
77.       //Ορίζει τιμή για το πεδίο εισαγωγής "Μετάβαση στη σελίδα" στον συνολικό αριθμό
σελίδων.
78.       document.getElementById('jumpPageInput').max = totalPages;
79.     }
80.
81.
82.     // Φίλτρα αναζήτησης
83.     document.getElementById('resourceSearch').addEventListener('input', function () {
84.       const search = this.value.toLowerCase().trim(); //Μέτατρέπουμε τα Γραμματα ολα σε
μικρά και αφαιρούμε τα κενά
85.       if (!search) { //Αν είναι τιμη κενη τα filteredItems είναι ολα τα Items
86.         filteredItems = allItems.slice();
87.       } else {
88.         // Διαχωρισμός φίλτρων με βάση το ;
89.         const filters = search.split(';').map(f => f.trim()); // Διαχωρισμός φίλτρων
90.
91.         filteredItems = allItems.filter(item => { //Αλλιώς τα filtered Items περνουν
την τιμή αναλογα με το τι πλήκτρολογουμε

```

```

92.                 const text = item.textContent.toLowerCase().replace(/_/g, ' '); //
Αντικατάσταση _ με κενά
93.                 return filters.every(filter => text.includes(filter)); // Επιστρέφουμε οτι
Indexes περιεχουν το ορισμα που δωσαμε στην μπάρα
94.             });
95.         }
96.         currentPage = 1; // μετά την αναζήτηση ειμαστε παντα στην 1η σελίδα
97.         updatePaginationForFilter(); // εκτελείτε η Function για τον υπολογισμο των total
pages μετά το φιλτραρισμα
98.     });
99.
100.
101.     //Μετά το φιλτραρισμα ενημερωνη το Total Page αναλογως με τον συνολικο αριθμο σελιδων
που θα προκυψουν
102.     function updatePaginationForFilter() {
103.         const totalPages2 = Math.ceil(filteredItems.length / pageSize);
104.         if (currentPage > totalPages2) {
105.             currentPage = totalPages2;
106.         }
107.         renderPage();
108.     }
109.
110.     // Μετάβαση σε συγκεκριμένη σελίδα
111.     document.getElementById('jumpPageButton').addEventListener('click', () => {
112.
113.         const targetPageInput = document.getElementById('jumpPageInput').value;
114.
115.         // 2. Προσθέτουμε έναν έλεγχο για επιτρεπτά ψηφία (π.χ. μόνο αριθμοί)
116.         if (!/^\\d+$/ .test(targetPageInput) || targetPageInput==0) {
117.             alert('Εισάγετε Μόνο Αριθμό Εκτός του 0');
118.             return;
119.         }
120.         const targetPage = parseInt(document.getElementById('jumpPageInput').value);
//μετατρέπει τον αριθμο που βαλαμε σε Int
121.         const totalPages = Math.ceil(filteredItems.length / pageSize); //Total Pages Ανάλογα
και με τα Filtered Items
122.         //Έλεγχος Αν η τιμή δεν είναι αιρθος και ειναι ειναι μεγαλυτερη η ιου του ενα και
μικροτερη η ιση του total page
123.
124.         if (!isNaN(targetPage) && targetPage >= 1 && targetPage <= totalPages) {
125.             currentPage = targetPage; //τοτε currnetPage θα παρει την τιμη του targetPage
126.             renderPage(); //εκτελείτε η function renderPage()
127.         } else {
128.             alert(`Υπάρχουν μόνο ${totalPages} Σελίδες`);
129.         }
130.     });
131.
132.     // render page Για Προηγουμενη Σελίδα
133.     document.getElementById('prevPage').addEventListener('click', () => {
134.         if (currentPage > 1) {
135.             currentPage--;
136.             renderPage();
137.         }
138.     });
139.     //renderPage για επομενη σελίδα
140.     document.getElementById('nextPage').addEventListener('click', () => {
141.         if (currentPage < Math.ceil(filteredItems.length / pageSize)) {
142.             currentPage++;
143.             renderPage();
144.         }
145.     });
146.
147.     // To render Page() χωρις φιλτρα η την σελιδοποιηση
148.     renderPage();
149. });
150. </script>
151. {% endblock %}
152.

```

Update Datasets Button

```

1. <!-- Update Datasets Button -->
2. <script src="https://code.jquery.com/jquery-3.6.0.min.js"></script>
3. <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.1.3/dist/js/bootstrap.bundle.min.js"></script>
4.
5. <a class="btn btn-primary" id="update-datasets-button" aria-label="Update Datasets">
6.   <i class="fa fa-sync"></i> Update Datasets</a>
7.
8. <div id="output-modal" class="modal fade" tabindex="-1" role="dialog">
9.   <div class="modal-dialog modal-lg" role="document">
10.    <div class="modal-content">
11.      <div class="modal-header">
12.        <h5 class="modal-title">Update Datasets</h5>
13.        <button type="button" class="btn btn-close" data-bs-dismiss="modal" aria-label="Close"></button>
14.      </div>
15.      <div class="modal-body">
16.        <pre id="output-content" style="white-space: pre-wrap; max-height: 400px; overflow-y: auto;"></pre>
17.      </div>
18.      <div class="modal-footer">
19.        <button type="button" class="btn btn-danger" id="terminate-button">Terminate</button>
20.        <button type="button" class="btn btn-secondary" data-bs-dismiss="modal">Close</button>
21.      </div>
22.    </div>
23.  </div>
24. </div>
25.
26. <script>
27. $(document).ready(function () { // Εκτελείται όταν το DOM είναι πλήρως φορτωμένο
28.   let RunningStatus = false; // Καταγραφή αν υπάρχει ενεργή διαδικασία
29.   let pollingInterval; // Μεταβλητή για την παρακολούθηση της κατάστασης
30.
31.   // Συνάρτηση για έλεγχο της κατάστασης
32.   function pollStatus() {
33.     $.ajax({
34.       url: "/waterhessaloniki/status", // Endpoint για έλεγχο κατάστασης
35.       method: "GET", // Μέθοδος HTTP για να πάρουμε δεδομένα
36.       success: function (data) { // Αν επιστραφεί επιτυχώς η κατάσταση
37.         if (data.status === "running") { // Αν η διαδικασία εκτελείται
38.           RunningStatus = true; // Σημειώνουμε ότι είναι ενεργή
39.           $('#update-datasets-button').prop('disabled', true); // Απενεργοποιούμε το κουμπί έναρξης
40.           $('#terminate-button').prop('disabled', false); // Ενεργοποιούμε το κουμπί τερματισμού
41.         } else { // Αν έχει ολοκληρωθεί
42.           RunningStatus = false; // Σημειώνουμε ότι η διαδικασία έχει σταματήσει
43.           $('#update-datasets-button').prop('disabled', false); // Ενεργοποιούμε το κουμπί έναρξης
44.           $('#terminate-button').prop('disabled', true); // Απενεργοποιούμε το κουμπί τερματισμού
45.
46.           // Εμφάνιση των αποτελεσμάτων της διαδικασίας
47.           $('#output-content').html(data.output);
48.
49.           // Κλείνουμε το modal μετά από 20 δευτερόλεπτα
50.           setTimeout(function () {
51.             $('#output-modal').modal('hide');
52.           }, 2000000000000000);
53.
54.           clearInterval(pollingInterval); // Σταματάμε την παρακολούθηση
55.         }
56.       },
57.       error: function () { // Αν υπάρξει σφάλμα
58.         $('#output-content').append('<br>Unable to fetch status. '); // Ενημερώνουμε τον χρήστη
59.       }
60.     });
61.   }
62.
63.   // Έναρξη ενημέρωσης
64.   $('#update-datasets-button').on('click', function (event) {
65.     event.preventDefault(); // Αποφυγή default συμπεριφοράς του κουμπιού
66.     $('#output-content').html(""); // Καθαρίζουμε το χώρο εμφάνισης αποτελεσμάτων
67.     $('#output-modal').modal('show'); // Εμφανίζουμε το modal
68.
69.     $.ajax({

```

```

70.     url: "/waterthessaloniki/start", // Endpoint για να ξεκινήσει η διαδικασία
71.     method: "POST", // Μέθοδος HTTP για να ξεκινήσει το script
72.     success: function (data) { // Αν ξεκινήσει επιτυχώς
73.         $('#output-content').append("Η ΕΡΓΑΣΙΑ ΞΕΚΙΝΗΣΕ ΝΑ ΕΚΤΕΛΕΙΤΕ ΜΕ PID: " + data.pid + "<br>"); //
Ενημερώνουμε για το PID
74.         pollingInterval = setInterval(pollStatus, 5000); // Επαναληπτικός έλεγχος κάθε 5 δευτερόλεπτα
75.         RunningStatus = true; // Σημειώνουμε ότι είναι ενεργή
76.     },
77.     error: function (xhr) { // Αν υπάρξει σφάλμα κατά την εκκίνηση
78.         $('#output-content').append("Error starting script: " + xhr.responseText + "<br>"); // Ενημερώνουμε τον χρήστη
79.     }
80. });
81. });
82.
83. // Τερματισμός της διαδικασίας
84. $('#terminate-button').on('click', function () {
85.     if (RunningStatus) { // Αν η διαδικασία είναι ενεργή
86.         $.ajax({
87.             url: "/waterthessaloniki/terminate", // Endpoint για τερματισμό
88.             method: "POST", // Μέθοδος HTTP για τον τερματισμό
89.             success: function (response) { // Αν τερματιστεί επιτυχώς
90.                 $('#output-content').append('<br>' + response.message); // Ενημερώνουμε τον χρήστη
91.                 RunningStatus = false; // Γυρνάμε το RunningStatus σε false Σημειώνουμε ότι σταμάτησε
92.                 clearInterval(pollingInterval); // Σταματάμε pollingInterval τον επαναληπτικό έλεγχο
93.             },
94.             error: function (xhr) { // Αν υπάρξει σφάλμα κατά τον τερματισμό
95.                 $('#output-content').append('<br>ΣΦΑΛΜΑ ' + xhr.responseText); // Ενημερώνουμε τον χρήστη
96.             }
97.         });
98.     } else {
99.         $('#output-content').append('<br>Δεν υπάρχει εκτέλουμένη εργασία του UPDATE DATASETS. '); // Ενημερώνουμε ότι
δεν υπάρχει κάτι να τερματίσουμε
100.     }
101. });
102. });
103. </script>
104.

```