



ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ
ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Από τους Χώρους Δεδομένων στη Διαμόρφωση
Πολιτικής: Ένα Πλαίσιο βασισμένο σε Ροές Εργασίας
για την Αξιοποίηση Ανοικτών και Εμπορικών
Οικοσυστημάτων Δεδομένων μέσω Πραγματικών
Αναλυτικών Σεναρίων»

"From Data Spaces to Policy Making: Workflow-Driven
Exploitation of Open and Commercial Data Ecosystems"

Φοιτήτρια:
Βουράκη Μαριάννα
Αρ. Μητρώου: 2019022

Επιβλέπων:
ΧΑΡΑΛΑΜΠΟΣ ΜΠΡΑΤΣΑΣ
Βαθμίδα: Καθηγητής

Θεσσαλονίκη, 2026

Τίτλος Δ.Ε.:

Από τους Χώρους Δεδομένων στη Διαμόρφωση Πολιτικής: Ένα Πλαίσιο βασισμένο σε Ροές Εργασίας για την Αξιοποίηση Ανοικτών και Εμπορικών Οικοσυστημάτων Δεδομένων μέσω Πραγματικών Αναλυτικών Σεναρίων.

Κωδικός Δ.Ε.: 26156

Ονοματεπώνυμο φοιτήτριας: ΒΟΥΡΑΚΗ ΜΑΡΙΑΝΝΑ

Ονοματεπώνυμο εισηγητή: ΧΑΡΑΛΑΜΠΟΣ ΜΠΡΑΤΣΑΣ

Ημερομηνία ανάληψης Δ.Ε.: 10-03-2026

Ημερομηνία περάτωσης Δ.Ε.: 26-05-2026

Βεβαιώνω ότι είμαι η 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 της φοιτήτριας Βουράκη Μαριάννα που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας 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

Στην οικογένειά μου, που ήταν πάντα δίπλα μου. Ιδιαίτερα στην αδερφή μου Βαλεντίνα, που με τη διαρκή της στήριξη και την παρουσία της, αποτελεί πρότυπο ζωής για μένα. Ένα ξεχωριστό ευχαριστώ οφείλω στον σύντροφό μου Θοδωρή, που ανέχτηκε τη γκρίνια μου με υπομονή και αγάπη σε όλη τη διάρκεια των σπουδών μου.

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά τη διάρκεια της πρακτικής μου άσκησης μέσω του Τμήματος Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος (ΔΙ.ΠΑ.Ε.). Στο πλαίσιο της άσκησης αυτής κλήθηκα να εργαστώ στην Αναπτυξιακή Μείζονος Αστικής Θεσσαλονίκης Α.Ε. - AOTA (MDAT), η οποία συμμετείχε στο ευρωπαϊκό ερευνητικό πρόγραμμα UPCAST (Universal Platform Components for Safe Fair Interoperable Data Exchange, Monetization and Trading), χρηματοδοτούμενο από το πρόγραμμα Horizon Europe της Ευρωπαϊκής Ένωσης (Grant Agreement No. 101092901).

Αντικείμενο της εργασίας μου ήταν η μελέτη και η πλήρης εφαρμογή ενός πλαισίου αξιοποίησης δεδομένων, βασισμένου σε αυτοματοποιημένες ροές επεξεργασίας (workflows), με σκοπό τη μετατροπή ανοικτών και εμπορικών δεδομένων σε δομημένους, συμμορφούμενους και επαναχρησιμοποιήσιμους δείκτες πολιτικής. Η μελέτη δεν περιορίστηκε σε θεωρητικό επίπεδο: η πλήρης εφαρμογή της πραγματοποιήθηκε μέσω των υφιστάμενων τεχνολογιών και υποδομών του UPCAST ecosystem συμπεριλαμβανομένου του Consumer Dashboard και αποτέλεσε μέρος της πραγματικής λειτουργίας του MDAT Pilot στη Θεσσαλονίκη.

Η επαφή με τις σημασιολογικές τεχνολογίες, τα Data Spaces και τα εργαλεία διακυβέρνησης δεδομένων αποτέλεσε για μένα πολύτιμη εμπειρία, η οποία με εισήγαγε στην ευρύτερη ευρωπαϊκή στρατηγική για τα δεδομένα και τις πρακτικές εφαρμογές της στον αστικό σχεδιασμό. Η παρούσα εργασία αντικατοπτρίζει αυτή τη διαδρομή: από την ανακάλυψη δεδομένων έως την παραγωγή δεικτών που χρησιμοποιήθηκαν για υποστήριξη αποφάσεων σε αστικό επίπεδο.

Περίληψη

Η ραγδαία αύξηση της παραγωγής δεδομένων από τεχνολογίες όπως το IoT, η Τεχνητή Νοημοσύνη και οι έξυπνες πόλεις έχει δημιουργήσει σύνθετα οικοσυστήματα Data Spaces και Data Marketplaces, με σημαντικό χάσμα μεταξύ της διαθεσιμότητας ετερογενών δεδομένων και της αξιοποίησής τους για τεκμηριωμένες αποφάσεις και πολιτικές. Η παρούσα διπλωματική εργασία εξετάζει τη σημασιολογική μοντελοποίηση ροών επεξεργασίας δεδομένων (Data Processing Workflows) με ενσωμάτωση των προτύπων Data Privacy Vocabulary (DPV 2.2) και Open Digital Rights Language (ODRL 2.2), στο πλαίσιο του ευρωπαϊκού έργου UPCAST και του MDAT Pilot (Αναπτυξιακή Μείζονος Αστικής Θεσσαλονίκης Α.Ε. - AOTA).

Η ερευνητική διαδικασία περιλάμβανε την αναζήτηση και επιλογή ανοικτών δεδομένων, τον σχεδιασμό αναλυτικών σεναρίων, την αυτοματοποίηση της επεξεργασίας μέσω Python και τη μετατροπή των αναλυτικών διαδικασιών σε σημασιολογικά τεκμηριωμένες ροές εργασίας. Αναπτύχθηκαν τρία σενάρια μελέτης σχετικά με δείκτες αστικού πρασίνου, ποιότητας αέρα και πληθυσμιακής έκθεσης, καθώς και κατανάλωση ενέργειας και εκπομπές CO₂. Για τις ανάγκες της υλοποίησης αναπτύχθηκε επίσης η προσαρμοσμένη οντολογία MDAT, με στόχο τη σημασιολογική περιγραφή των διαδικασιών, των εννοιών και των σχέσεων που εμφανίζονται στα αναλυτικά σενάρια και στις αντίστοιχες ροές επεξεργασίας δεδομένων. Κάθε βήμα αντιστοιχίστηκε σε κλάσεις του DPV και πολιτικές του ODRL, ενώ εισήχθησαν εξειδικευμένες έννοιες (χώρος ονομάτων: mdat:) για την περιγραφή των δεδομένων πέρα από τους γενικούς όρους των δύο προτύπων, υποστηρίζοντας τη διαλειτουργικότητα και την επεκτασιμότητα.

Βασική συνεισφορά της εργασίας αποτελεί η μετατροπή συμβατικών Python scripts σε δομημένα, επαναχρησιμοποιήσιμα workflows με επίγνωση πολιτικής. Μέσω της οντολογίας MDAT, κάθε βήμα μοντελοποιήθηκε ως `dpv:Process` συνδεδεμένο με πολιτικές ODRL, επιτρέποντας συμμόρφωση εκ σχεδιασμού, ιχνηλασιμότητα και αυτοματοποιημένη επιβολή περιορισμών. Κεντρικό εύρημα είναι ότι η σημασιολογική αντιστοίχιση μπορεί να λειτουργήσει ως ενεργός μηχανισμός διαμόρφωσης ροής και όχι ως παθητικό μεταδεδομένο, σε ελεγχόμενο πιλοτικό περιβάλλον. Η προσέγγιση ενισχύει την αναπαραγωγικότητα, τη διαφάνεια και τη συμμόρφωση μέσω κανόνων, παρέχοντας επεκτάσιμο πλαίσιο για έξυπνες πόλεις και Χώρους Δεδομένων.

Λέξεις-κλειδιά:

Data Processing Workflows, Semantic Modeling, DPV, ODRL, Data Spaces, Smart Cities, Compliance-by-Design, MDAT Ontology, UPCAST, Policy-aware Analytics

Abstract

The rapid increase in data production from technologies such as IoT, Artificial Intelligence, and smart cities has created complex Data Spaces and Data Marketplaces ecosystems, resulting in a significant gap between the availability of heterogeneous data and its utilization for informed decisions and policies. This thesis examines the semantic modeling of Data Processing Workflows by integrating the Data Privacy Vocabulary (DPV 2.2) and Open Digital Rights Language (ODRL 2.2) standards, within the framework of the European project UPGAST and the MDAT Pilot (Major Development Agency Thessaloniki S.A. - L.G.A.).

The research methodology involved the discovery and selection of open data, the design of analytical scenarios, the automation of processing via Python, and the transformation of analytical procedures into semantically documented workflows. Three case studies were developed focusing on urban green indicators, air quality and population exposure, as well as energy consumption and CO₂ emissions. For implementation purposes, the customized MDAT ontology was also developed, aiming at the semantic description of processes, concepts, and relationships appearing in the analytical scenarios and their respective data processing workflows. Each step was mapped to DPV classes and ODRL policies, while specialized concepts (namespace: mdat:) were introduced to describe data beyond the general terms of the two standards, thereby supporting interoperability and scalability.

The main contribution of this work lies in the transformation of conventional Python scripts into structured, reusable, and policy-aware workflows. Through the MDAT ontology, each step was modeled as a `dpv:Process` linked to ODRL policies, enabling compliance-by-design, traceability, and automated enforcement of restrictions. A key finding is that semantic mapping can function as an active workflow configuration mechanism rather than passive metadata within a controlled pilot environment. This approach enhances reproducibility, transparency, and rule-based compliance, providing a scalable framework for smart cities and Data Spaces.

Keywords:

Data Processing Workflows, Semantic Modeling, DPV, ODRL, Data Spaces, Smart Cities, Compliance-by-Design, MDAT Ontology, UPGAST, Policy-aware Analytics

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στην ομάδα της Open Knowledge Foundation Greece (OKFN Greece) και στην Αναπτυξιακή Μείζονος Αστικής Θεσσαλονίκης (MDAT) για την ευκαιρία που μου δόθηκε να εργαστώ στο πλαίσιο του ευρωπαϊκού προγράμματος UPCAST. Η εμπειρία αυτή αποτέλεσε σπάνια ευκαιρία για εφαρμογή ακαδημαϊκών γνώσεων σε πραγματικό ερευνητικό και τεχνολογικό περιβάλλον.

Ιδιαίτερες ευχαριστίες οφείλω στον επιβλέποντα καθηγητή μου για την καθοδήγηση και υποστήριξη, καθώς και στους συναδέλφους του UPCAST project για τη συνεργασία. Τέλος, ευχαριστώ θερμά την οικογένειά μου για την αμέριστη συμπαράστασή της.

ΠΕΡΙΕΧΟΜΕΝΑ

Πρόλογος.....	4
Περίληψη.....	5
Abstract	6
Ευχαριστίες	7
Συντομογραφίες.....	12
Κεφάλαιο 1ο: Εισαγωγή.....	13
1.1 Η Εποχή των Δεδομένων - Πλαίσιο και Κίνητρα	13
1.2 Το Πρόβλημα: Από τα Ακατέργαστα Δεδομένα στην Πολιτική Γνώση.....	13
1.3 Η Προτεινόμενη Λύση: Workflow-Driven Αξιοποίηση	13
1.4 Στόχοι, Μεθοδολογία και Δομή Εργασίας	14
Κεφάλαιο 2ο: Θεωρητικό Υπόβαθρο και Σχετική Βιβλιογραφία	15
2.1 Data Spaces	15
2.2 Η Άνοδος των Data Spaces και Data Marketplaces	15
2.2.1 Open Data Spaces.....	15
2.2.2 Data Marketplaces	16
2.2.3 Το Πλαίσιο Αδειοδότησης (Licensing Landscape)	16
2.3 Παραγωγοί Δεδομένων και Κατηγορίες Datasets	17
2.3.1 Αστικά και Δημόσια Οικοσυστήματα	17
2.3.2 Ιδιωτικός Τομέας και IoT	17
2.3.3 Κατηγορίες Δεδομένων	18
2.4 Ο Σημασιολογικός Ιστός (Semantic Web)	18
2.4.1 Resource Description Framework (RDF).....	19
2.4.2 SPARQL - Η Γλώσσα Ερωτημάτων του RDF	19
2.4.3 OWL και RDFS - Γλώσσες Οντολογιών.....	19
2.5 Linked Data	19
2.6 Οντολογίες (Ontologies)	20
2.6.1 Τι είναι Οντολογία;	20
2.6.2 Γιατί Χρειαζόμαστε Οντολογίες στα Data Spaces;	21
2.6.3 Η Οντολογία MDAT	21
2.7 Data Privacy Vocabulary (DPV)	22
2.7.1 Τι είναι το DPV;	22
2.7.2 DPV στα Workflows Επεξεργασίας Δεδομένων.....	23

2.8 Open Digital Rights Language (ODRL).....	23
2.8.1 Τι είναι το ODRL;	23
2.8.2 ODRL Policy Παράδειγμα	23
2.8.3 Συνδυασμός ODRL + DPV για Compliance-by-Design.....	23
2.9 Data Processing Workflows: Έννοια και Αξία.....	24
2.9.1 Τι είναι Data Processing Workflow;	24
2.9.2 Η Αξία της Επαναχρησιμοποίησης Workflows	24
Σχετική Βιβλιογραφία	24
Κεφάλαιο 3ο: Το Οικοσύστημα UPCAST	25
3.1 Το Thessaloniki Data Space (TDS).....	25
3.2 Τα Εργαλεία UPCAST	25
3.3 Consumer Dashboard	26
3.4 Το Πλαίσιο Data Processing Workflow (DPW) εντός του UPCAST	27
3.4.1 Αρχιτεκτονική και Κύκλος Ζωής DPW	27
Στάδιο 1: Discovery and Design	28
Στάδιο 2: Semantic Alignment.....	28
Στάδιο 3: Policy Negotiation and Re-engineering.....	28
Στάδιο 4: Execution and Monitoring.....	29
Στάδιο 5: Refinement and Reuse.....	29
3.5 Από το Οικοσύστημα στην Πράξη: Εφαρμογή σε Πραγματικά Σενάρια	30
Κεφάλαιο 4ο: Μελέτες Περίπτωσης-Workflow-Driven Ανάλυση	31
4.1 Σενάριο 1: Δείκτες Αστικού Πρασίνου	31
4.1.1 Πλαίσιο και Ερευνητικός Στόχος.....	31
4.1.2 Πηγές Δεδομένων.....	31
4.1.3 Ρόλοι Οντολογίας.....	32
4.1.4 Σημασιολογική Αντιστοίχιση Διεργασιών	32
4.1.5 Ροή Επεξεργασίας Δεδομένων	32
4.1.6 Δομή Παραδοτέων και Αποτελέσματα Εκτέλεσης	34
4.1.7 Συμπεράσματα του Σεναρίου	37
4.2 Σενάριο 2: Δείκτες Ποιότητας Αέρα και Πληθυσμιακής Έκθεσης.....	38
4.2.1 Πλαίσιο και Ερευνητικός Στόχος.....	38
4.2.2 Πηγές Δεδομένων.....	38
4.2.3 Ρόλοι Οντολογίας.....	39
4.2.4 Διαπραγμάτευση και Σταδιακή Πρόσβαση	39
4.2.5 Αυτόματη Επανασχεδίαση Workflow.....	40

4.2.6 Σημασιολογική Αντιστοίχιση Διεργασιών	40
4.2.7 Ροή Επεξεργασίας Δεδομένων	41
4.2.8 Δομή Παραδοτέων και Αποτελέσματα Εκτέλεσης	42
4.2.9 Συμπεράσματα του Σεναρίου	49
4.3 Σενάριο 3: Ενεργειακές και Περιβαλλοντικές Μετρήσεις	50
4.3.1 Πλαίσιο και Ερευνητικός Στόχος	50
4.3.2 Πηγές Δεδομένων	50
4.3.3 Ρόλοι Οντολογίας	51
4.3.4 Σημασιολογική Αντιστοίχιση Διεργασιών	51
4.3.5 Ροή Επεξεργασίας Δεδομένων	52
4.3.6 Δομή Παραδοτέων και Αποτελέσματα Εκτέλεσης	53
4.3.7 Συμπεράσματα του Σεναρίου	57
Κεφάλαιο 5ο: Αποτελέσματα και Συζήτηση	58
5.1 Συγκριτική Σύθεση των Τριών Σεναρίων	58
5.2 Compliance-by-Design: Θεωρητική Εφαρμογή	59
5.3 Επαναχρησιμοποίηση και Επεκτασιμότητα	60
5.4 Συμπληρωματικότητα των Σεναρίων	60
5.5 Περιορισμοί	60
Κεφάλαιο 6ο: Συμπεράσματα και Μελλοντική Εργασία	62
6.1 Σύνοψη της Συνεισφοράς	62
6.2 Θεωρητικές Συνέπειες	63
6.3 Τοποθέτηση στο Ευρωπαϊκό Πλαίσιο	63
6.4 Μελλοντικές Κατευθύνσεις	63
6.5 Επίλογος	64
Βιβλιογραφία	65
Παραρτήματα Κεφαλαίων	66
Παράρτημα Α: Επέκταση Κεφαλαίου 1 - Το Παγκόσμιο Τοπίο Παραγωγής Δεδομένων	67
A.1 Στατιστικά Παραγωγής Δεδομένων	67
A.2 Γεωγραφική Κατανομή Data Centers	67
A.3 Ελληνικές Πύλες Ανοικτών Δεδομένων	67
Παράρτημα Β: Επέκταση Κεφαλαίου 2 - Τεχνικές Λεπτομέρειες Σημασιολογικών Τεχνολογιών	68
B.1 RDF Turtle Παράδειγμα για MDAT Dataset	68
B.2 ODRL Policy Παράδειγμα για Restricted Dataset	68
B.3 Σύγκριση Λειτουργιών	69
Παράρτημα Γ: Επέκταση Κεφαλαίου 3 - Τεχνικές Λεπτομέρειες DPW	70

Γ.1 Αρχιτεκτονική Nextflow	70
Γ.2 Συνδυασμός ODRL + DPV στη Πράξη: Βήμα-Βήμα.....	70
Γ.3 Σύγκριση Σεναρίων: Χαρακτηριστικά DPW	70
Παράρτημα Δ: Επέκταση Κεφαλαίου 5 - Αναλυτικά Αποτελέσματα.....	72
Δ.1 Αποτελέσματα Σεναρίου 2: Ρύπανση ανά Δημοτική Κοινότητα	72
Δ.2 Τύπος Υπολογισμού CO ₂ Εκπομπών (Σενάριο 3)	72
Δ.3 Πηγές Δεδομένων TDS.....	73
Παράρτημα Ε: Γλωσσάριο Τεχνικών Όρων.....	74
Παράρτημα ΣΤ: Πλατφόρμες Δεδομένων και Εργαλεία	75
ΣΤ.1 Ελληνικές Πύλες Ανοικτών Δεδομένων.....	75
ΣΤ.2 Εμπορικές Πλατφόρμες Δεδομένων.....	75
ΣΤ.3 Εργαλεία που Χρησιμοποιήθηκαν.....	75

Συντομογραφίες

Συντομογραφία	Πλήρης Ονομασία
ABOVO	ICT Abovo Consumer Dashboard Platform
API	Application Programming Interface
CC-BY	Creative Commons Attribution
DaaS	Data-as-a-Service
Δ.Ε.	Διπλωματική Εργασία
ΔΙ.ΠΑ.Ε.	Διεθνές Πανεπιστήμιο Ελλάδος
DPV	Data Privacy Vocabulary
DPW	Data Processing Workflow
GDPR	General Data Protection Regulation
IoT	Internet of Things
JSON-LD	JavaScript Object Notation for Linked Data
MDAT	Αναπτυξιακή Μείζονος Αστικής Θεσσαλονίκης / Custom Ontology Namespace
ODbL	Open Database License
ODRL	Open Digital Rights Language
OKFN	Open Knowledge Foundation
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	RDF Schema
SPARQL	SPARQL Protocol and RDF Query Language
TDS	Thessaloniki Data Space
TURTLE	Terse RDF Triple Language
UPCAST	Universal Platform Components for Safe Fair Interoperable Data Exchange, Monetization and Trading
URI	Uniform Resource Identifier
W3C	World Wide Web Consortium

Κεφάλαιο 1ο: Εισαγωγή

1.1 Η Εποχή των Δεδομένων - Πλαίσιο και Κίνητρα

Ζούμε σε μια εποχή όπου τα δεδομένα έχουν καταστεί ο πιο πολύτιμος πόρος του ψηφιακού κόσμου. Η ραγδαία αύξηση της παραγωγής τους αποτελεί αναμφισβήτητο γεγονός: σύμφωνα με τη Statista, ο παγκόσμιος όγκος δεδομένων που παρήχθησαν το 2025 αναμένεται να φτάσει τα 181 zettabytes, με προβλέψεις που ανεβάζουν τον αριθμό αυτό στα 394 zettabytes έως το 2028 [1]. Η εκθετική αυτή άνοδος οφείλεται σε έναν συνδυασμό αλληλένδετων παραγόντων: την καθολική υιοθέτηση τεχνολογιών τεχνητής νοημοσύνης (AI) και υπολογιστικού νέφους (Cloud), την ταχεία εξάπλωση του Internet of Things (IoT) με αναμενόμενο αριθμό 40,6 δισεκατομμυρίων συνδεδεμένων συσκευών έως το 2034, και τις επενδύσεις στον ψηφιακό μετασχηματισμό που αναμένεται να αγγίξουν τα 3,88 τρισεκατομμύρια δολάρια μέχρι το 2028.

Η υλική υποδομή αυτής της εξέλιξης αποτυπώνεται σαφώς στον αριθμό των data centers: σύμφωνα με το Data Center Map, καταγράφονται σήμερα 11.028 ενεργές εγκαταστάσεις παγκοσμίως, με τις ΗΠΑ να ηγούνται με 4.036, το Ηνωμένο Βασίλειο με 511, τη Γερμανία με 507, και την Ελλάδα να συμμετέχει με 25 εγκαταστάσεις [2]. Αυτή η μαζική παραγωγή δεδομένων από IoT, έξυπνες πόλεις, δημόσιους φορείς και επιχειρήσεις τροφοδοτεί την ανάγκη για νέα οικοσυστήματα διαχείρισης: τα Data Spaces και τα Data Marketplaces.

1.2 Το Πρόβλημα: Από τα Ακατέργαστα Δεδομένα στην Πολιτική Γνώση

Παρά τη ραγδαία αύξηση της παραγωγής δεδομένων, το χάσμα μεταξύ της διαθεσιμότητας ακατέργαστων δεδομένων και της πρακτικής αξιοποίησής τους για την υποστήριξη αποφάσεων παραμένει μεγάλο [3]. Τα βασικά εμπόδια είναι:

- Κατακερματισμός πηγών: Τα δεδομένα βρίσκονται σε ετερογενείς πλατφόρμες, χωρίς ενοποιημένη αναζήτηση [4].
- Ετερογένεια μορφών: Διαφορετικά σχήματα, πρότυπα και γλώσσες κωδικοποίησης
- Ποικιλομορφία αδειοδότησης: Ασυνεπείς όροι χρήσης που περιορίζουν την επαναχρησιμοποίηση [5], [6].
- Χειρωνακτικές pipelines: Η απουσία αυτοματοποίησης καθιστά την ανάλυση χρονοβόρα και επιρρεπή σε σφάλματα.
- Έλλειψη νομικής διαφάνειας: Χωρίς δομημένη παρακολούθηση, η συμμόρφωση με GDPR και άλλες υποχρεώσεις είναι πρακτικά αδύνατη [6], [7].

Η αντιμετώπιση αυτών των εμποδίων απαιτεί ένα ολοκληρωμένο πλαίσιο που να καλύπτει ολόκληρο τον κύκλο ζωής των δεδομένων: από την ανακάλυψη και την απόκτηση, έως την επεξεργασία, την ανάλυση και την οπτικοποίηση αποτελεσμάτων με ενσωματωμένη σε κάθε βήμα τη νομική και ηθική συμμόρφωση [8].

1.3 Η Προτεινόμενη Λύση: Workflow-Driven Αξιοποίηση

Η παρούσα εργασία παρουσιάζει και εφαρμόζει το πλαίσιο Data Processing Workflow (DPW), μια δομημένη, policy-aware προσέγγιση που τυποποιεί ολόκληρη τη διαδρομή από την ανακάλυψη δεδομένων έως την παραγωγή αναπαραγωγίμων, compliance-aware δεικτών πολιτικής. Το πλαίσιο αξιοποιεί τεχνολογίες του Σημασιολογικού Ιστού (Semantic Web) συμπεριλαμβανομένων οντολογιών, RDF, ODRL και DPV [7], [8], [9] ενσωματώνοντας πολιτικές χρήσης και περιορισμούς απορρήτου απευθείας στις εκτελέσιμες αναλυτικές pipelines.

Στο πλαίσιο αυτό, η παρούσα εργασία διαρθρώνεται γύρω από τρία ερευνητικά ερωτήματα:

- RQ1: Πώς μπορεί η σημασιολογική περιγραφή δεδομένων να μετατραπεί από παθητικό μεταδεδομένο σε ενεργό μηχανισμό διαμόρφωσης ροών εργασίας;
- RQ2: Σε ποιο βαθμό μπορεί η αρχή του Compliance-by-Design να αυτοματοποιήσει τη νομική και ηθική συμμόρφωση χωρίς ανθρώπινη παρέμβαση, μέσω της δυναμικής επανασχεδίασης ροών εργασίας;
- RQ3: Σε ποιον βαθμό το προτεινόμενο πλαίσιο είναι επεκτάσιμο και επαναχρησιμοποιήσιμο σε διαφορετικούς τύπους αναλυτικών σεναρίων (ανοικτά δεδομένα, δεδομένα περιορισμένης πρόσβασης, διαχρονικές αναλύσεις);

1.4 Στόχοι, Μεθοδολογία και Δομή Εργασίας

Οι κύριοι στόχοι της εργασίας είναι:

1. Ανάπτυξη ενός DPW μοντέλου για δομημένη διαχείριση κύκλου ζωής ετερογενών δεδομένων εντός Data Space
2. Υλοποίηση Compliance-by-Design μηχανισμού που ενσωματώνει ODRL/DPV απευθείας στη δομή workflow
3. Παρουσίαση της μεθοδολογίας μέσω τριών πραγματικών σεναρίων ανάλυσης πάνω σε δημόσια δεδομένα της Θεσσαλονίκης
4. Υποστήριξη λήψης αποφάσεων αστικής πολιτικής μέσω τεκμηριωμένων, νομικά συμμορφούμενων αποτελεσμάτων

Η μεθοδολογία οργανώνεται σε τέσσερα στάδια: (α) ανασκόπηση βιβλιογραφίας και επιλογή ανοικτών δεδομένων, (β) σχεδιασμός τριών αναλυτικών σεναρίων, (γ) αυτοματοποίηση εκτέλεσης μέσω Python και Nextflow, και (δ) σημασιολογική μοντελοποίηση μέσω MDAT οντολογίας [10].

Η εργασία οργανώνεται ως εξής: Το Κεφάλαιο 2 αναλύει το θεωρητικό υπόβαθρο. Το Κεφάλαιο 3 περιγράφει το οικοσύστημα UPCAST. Το Κεφάλαιο 4 αναλύει τα τρία σενάρια. Τα Κεφάλαια 5 και 6 συζητούν αποτελέσματα και συμπεράσματα. Στα Παραρτήματα παρέχονται επεκτάσεις κεφαλαίων και τεχνικές λεπτομέρειες.

Κεφάλαιο 2ο: Θεωρητικό Υπόβαθρο και Σχετική Βιβλιογραφία

2.1 Data Spaces

Τα Data Spaces χαρακτηρίζονται ως αποκεντρωμένα, οικοσυστήματα που διασφαλίζουν την κυριαρχία των δεδομένων [11] , ένα όραμα που επισημοποιήθηκε από την Ευρωπαϊκή Στρατηγική για τα Δεδομένα και υλοποιήθηκε συγκεκριμένα μέσω πρωτοβουλιών όπως ο Χώρος Δεδομένων της Θεσσαλονίκης (TDS) στο πλαίσιο του UPCAST [3]. Η διαλειτουργικότητα σε τέτοια οικοσυστήματα βασίζεται σε τεχνολογίες Σημασιολογικού Ιστού.

2.2 Η Άνοδος των Data Spaces και Data Marketplaces

Η άνοδος των ψηφιακών αποθετηρίων δεδομένων έχει δημιουργήσει ένα νέο επιστημονικό και επιχειρηματικό περιβάλλον. Η διάθεση και η προσβασιμότητα των δεδομένων δημιουργούν σημαντικές ευκαιρίες και ανοίγουν νέους ορίζοντες για την έρευνα, την παρακολούθηση διαφόρων οικοσυστημάτων και την εξαγωγή χρήσιμων συμπερασμάτων μέσω της ανάλυσης και αξιοποίησής τους. Μέσα από αυτή τη διαδικασία καθίσταται δυνατή η υποστήριξη τεκμηριωμένων αποφάσεων και η βελτιστοποίηση τόσο της καθημερινής όσο και της ψηφιακής μας ζωής [12].

Η διάθεση των δεδομένων οργανώνεται κυρίως σε δύο βασικά μοντέλα. Από τη μία πλευρά, υπάρχουν τα Open Data Spaces, τα οποία παρέχουν ελεύθερη πρόσβαση σε δεδομένα χωρίς οικονομικό κόστος, προωθώντας την ανοικτή επιστήμη και τη διαφάνεια. Από την άλλη πλευρά, υπάρχουν τα Data Marketplaces, όπου τα δεδομένα διατίθενται με οικονομικό αντίτιμο, επιτρέποντας την εμπορική αξιοποίηση και τη δημιουργία νέων επιχειρηματικών μοντέλων βασισμένων στη διαχείριση και ανταλλαγή δεδομένων.

2.2.1 Open Data Spaces

Τα Open Data Spaces αποτελούν πλατφόρμες και οικοσυστήματα στα οποία δεδομένα διατίθενται δημόσια και μπορούν να χρησιμοποιηθούν ελεύθερα για έρευνα, ανάλυση και ανάπτυξη εφαρμογών. Τα δεδομένα αυτά προέρχονται συχνά από δημόσιους οργανισμούς, ερευνητικά ιδρύματα ή διεθνείς οργανισμούς και συνοδεύονται από open licenses, επιτρέποντας την επαναχρησιμοποίησή τους.

Ενδεικτικά παραδείγματα Open Data Spaces περιλαμβάνουν:

- Wikidata - Ένα από τα μεγαλύτερα ανοικτά knowledge graphs, το οποίο παρέχει δομημένα δεδομένα που μπορούν να χρησιμοποιηθούν για εφαρμογές τεχνητής νοημοσύνης, linked data και semantic web.
- EU Open Data Portal (data.europa.eu) – Η κεντρική ευρωπαϊκή πύλη ανοικτών δεδομένων, η οποία συγκεντρώνει χιλιάδες δεδομένα από οργανισμούς της Ευρωπαϊκής Ένωσης και κράτη μέλη, καλύπτοντας τομείς όπως το περιβάλλον, η οικονομία και οι μεταφορές.
- Data.gov - Η επίσημη πλατφόρμα ανοικτών δεδομένων της κυβέρνησης των Ηνωμένων Πολιτειών, η οποία παρέχει πρόσβαση σε χιλιάδες δεδομένα από δημόσιους φορείς για χρήση σε έρευνα, επιχειρηματικές εφαρμογές και ανάπτυξη ψηφιακών υπηρεσιών.
- Thessaloniki Data Space (TDS) - Πλατφόρμα ανοικτών δεδομένων για την πόλη της Θεσσαλονίκης, η οποία περιλαμβάνει δεδομένα που σχετίζονται με την ατμοσφαιρική ρύπανση, την υγεία, την κινητικότητα και άλλους αστικούς [13].

- OpenAIRE Research Graph - Ευρωπαϊκή υποδομή ανοικτής επιστήμης που παρέχει πρόσβαση σε επιστημονικά δεδομένα, publications και research datasets που προέρχονται από ερευνητικά έργα και πανεπιστήμια.

Τα παραπάνω παραδείγματα δείχνουν πως τα Open Data Spaces λειτουργούν ως ανοικτά οικοσυστήματα δεδομένων, τα οποία διευκολύνουν την ανακάλυψη, την πρόσβαση και την αξιοποίηση δεδομένων από ερευνητές, επιχειρήσεις και δημόσιους οργανισμούς, συμβάλλοντας στην ανάπτυξη καινοτόμων εφαρμογών και στη λήψη τεκμηριωμένων αποφάσεων.

2.2.2 Data Marketplaces

Η εμπορική πρόσβαση σε δεδομένα (Commercial Data Access) αφορά πλατφόρμες και αγορές δεδομένων στις οποίες οργανισμοί, επιχειρήσεις ή ερευνητές μπορούν να αποκτήσουν πρόσβαση σε δεδομένα έναντι οικονομικού ανταλλάγματος. Τα δεδομένα αυτά διατίθενται συνήθως μέσω Data Marketplaces, API services ή μοντέλων Data-as-a-Service (DaaS) στο πλαίσιο ολοκληρωμένων αρχιτεκτονικών [3], [11].

Ενδεικτικά παραδείγματα περιλαμβάνουν:

- Nokia Data Marketplace (NDM) - Πλατφόρμα ανταλλαγής δεδομένων που επιτρέπει την ασφαλή πρόσβαση σε δεδομένα για εφαρμογές τεχνητής νοημοσύνης και analytics, κυρίως στον τομέα των τηλεπικοινωνιών και των έξυπνων πόλεων.
- AWS Data Exchange - Marketplace της Amazon όπου οργανισμοί μπορούν να αγοράσουν ή να αποκτήσουν πρόσβαση μέσω συνδρομής σε δεδομένα από διάφορους παρόχους, όπως οικονομικά δεδομένα, γεωχωρικά δεδομένα ή δεδομένα αγοράς.
- Snowflake Data Marketplace - Πλατφόρμα που επιτρέπει την πρόσβαση σε εμπορικά δεδομένα απευθείας μέσα από το cloud περιβάλλον της Snowflake χωρίς την ανάγκη μεταφοράς των δεδομένων.
- Dawex Data Exchange Platform - Πλατφόρμα ανταλλαγής δεδομένων που χρησιμοποιείται από επιχειρήσεις και οργανισμούς για την εμπορική διάθεση και διαχείριση δεδομένων σε διάφορους κλάδους.
- Google Cloud Marketplace Datasets - Υπηρεσία μέσω της οποίας μπορούν να αποκτηθούν δεδομένα για analytics και machine learning εφαρμογές, συχνά μέσω APIs ή cloud storage.

Τα παραπάνω παραδείγματα δείχνουν πως τα commercial data ecosystems επιτρέπουν τη δημιουργία νέων επιχειρηματικών μοντέλων γύρω από τα δεδομένα, διευκολύνοντας την αξιοποίηση εξειδικευμένων datasets σε τομείς όπως η τεχνητή νοημοσύνη, οι τηλεπικοινωνίες, η οικονομία και οι έξυπνες πόλεις.

2.2.3 Το Πλαίσιο Αδειοδότησης (Licensing Landscape)

Η διάθεση και η επαναχρησιμοποίηση των δεδομένων μέσα στα Data Spaces και τα Data Marketplaces καθορίζεται σε μεγάλο βαθμό από το πλαίσιο αδειοδότησης (licensing landscape) που συνοδεύει κάθε dataset. Οι άδειες χρήσης καθορίζουν τους όρους πρόσβασης, χρήσης, διαμοιρασμού και επαναχρησιμοποίησης των δεδομένων, διασφαλίζοντας τόσο την προστασία των παρόχων δεδομένων όσο και τη σωστή αξιοποίησή τους από τρίτους.

Μία από τις πιο διαδεδομένες κατηγορίες είναι οι open licenses, οι οποίες επιτρέπουν την ελεύθερη πρόσβαση και χρήση των δεδομένων με ελάχιστους περιορισμούς. Παραδείγματα αποτελούν οι άδειες Open Database License (ODbL), Creative Commons Attribution (CC-BY) και Creative Commons Zero (CC0). Οι άδειες αυτές χρησιμοποιούνται ευρέως σε πλατφόρμες ανοικτών δεδομένων, επιτρέποντας την επαναχρησιμοποίηση των datasets για ερευνητικούς,

εκπαιδευτικούς ή ακόμη και εμπορικούς σκοπούς, υπό την προϋπόθεση ότι τηρούνται συγκεκριμένοι όροι, όπως η αναφορά της πηγής.

Παράλληλα, υπάρχουν και *restricted licenses*, οι οποίες επιβάλλουν συγκεκριμένους περιορισμούς στη χρήση των δεδομένων. Σε αυτή την περίπτωση, τα δεδομένα μπορεί να είναι διαθέσιμα μόνο για συγκεκριμένες κατηγορίες χρηστών (π.χ. ερευνητές ή οργανισμούς), να απαιτούν διαδικασία έγκρισης για την πρόσβαση ή να απαγορεύουν την εμπορική αξιοποίησή τους. Τέτοιες άδειες χρησιμοποιούνται συχνά σε *datasets* που περιέχουν ευαίσθητα δεδομένα, όπως ιατρικά ή προσωπικά δεδομένα.

Επιπλέον, σε πολλά *Data Marketplaces* εφαρμόζονται εμπορικές συμφωνίες, μέσω των οποίων οι χρήστες αποκτούν πρόσβαση σε δεδομένα έναντι οικονομικού ανταλλάγματος. Τα συμβόλαια αυτά καθορίζουν τους όρους χρήσης, τη διάρκεια πρόσβασης, τα δικαιώματα αναδιανομής και τα επίπεδα ασφάλειας των δεδομένων. Τα εμπορικά μοντέλα δεδομένων επιτρέπουν σε οργανισμούς και επιχειρήσεις να δημιουργούν νέα επιχειρηματικά μοντέλα βασισμένα στην αξιοποίηση και ανταλλαγή δεδομένων[3].

Τέλος, ένα κρίσιμο στοιχείο για τη σωστή λειτουργία των οικοσυστημάτων δεδομένων είναι η διακυβέρνηση και συμμόρφωση δεδομένων, δηλαδή το σύνολο των πολιτικών, διαδικασιών και κανονισμών που διασφαλίζουν τη νόμιμη και υπεύθυνη χρήση των δεδομένων. Οι μηχανισμοί αυτοί περιλαμβάνουν τη συμμόρφωση με κανονιστικά πλαίσια, όπως ο GDPR [6], την εφαρμογή πολιτικών ασφάλειας και ιδιωτικότητας μέσω προτύπων λεξιλογίων[7], καθώς και τη διασφάλιση της διαφάνειας στη διαχείριση και επεξεργασία των δεδομένων.

Συνολικά, το πλαίσιο αδειοδότησης αποτελεί βασικό παράγοντα για τη λειτουργία των *Data Spaces*, καθώς καθορίζει τον τρόπο με τον οποίο τα δεδομένα μπορούν να διαμοιραστούν, να επαναχρησιμοποιηθούν και να αξιοποιηθούν σε ερευνητικές, επιχειρηματικές και κοινωνικές εφαρμογές.

2.3 Παραγωγοί Δεδομένων και Κατηγορίες Datasets

2.3.1 Αστικά και Δημόσια Οικοσυστήματα

Σημαντικός όγκος δεδομένων παράγεται από αστικά και δημόσια οικοσυστήματα, στα οποία περιλαμβάνονται δήμοι, περιφέρειες και δημόσιοι οργανισμοί. Οι φορείς αυτοί συλλέγουν και δημοσιεύουν δεδομένα που σχετίζονται με τη λειτουργία των πόλεων, τη δημόσια διοίκηση και την παροχή υπηρεσιών προς τους πολίτες, όπως αποτυπώνεται χαρακτηριστικά στην περίπτωση του *Thessaloniki Data Space* [13].

Οι δήμοι παράγουν δεδομένα που αφορούν την κυκλοφορία, τις μεταφορές, τη διαχείριση απορριμμάτων, τη χρήση δημόσιων χώρων και την ποιότητα του περιβάλλοντος. Τα δεδομένα αυτά συχνά δημοσιεύονται μέσω ανοιχτές πύλες δεδομένων στο πλαίσιο των πολιτικών ανοικτών δεδομένων και της διαφάνειας. Οι περιφέρειες συγκεντρώνουν δεδομένα μεγαλύτερης γεωγραφικής κλίμακας, τα οποία σχετίζονται με την περιφερειακή ανάπτυξη, την οικονομία, τις υποδομές και τον χωροταξικό σχεδιασμό.

Παράλληλα, δημόσιοι οργανισμοί όπως εθνικές στατιστικές υπηρεσίες (ΕΛΣΤΑΤ), οργανισμοί περιβάλλοντος και μεταφορών ή ενεργειακές αρχές παράγουν και διαθέτουν δεδομένα που χρησιμοποιούνται για την ανάλυση κοινωνικών, οικονομικών, πληθυσμιακών και περιβαλλοντικών δεικτών.

2.3.2 Ιδιωτικός Τομέας και IoT

Εκτός από τον δημόσιο τομέα, σημαντικός όγκος δεδομένων παράγεται και από τον ιδιωτικό τομέα και τις τεχνολογικές υποδομές που σχετίζονται με το IoT. Οι εταιρείες τηλεπικοινωνιών,

οι πάροχοι ενέργειας και οι επιχειρήσεις που διαχειρίζονται δίκτυα αισθητήρων συλλέγουν συνεχώς δεδομένα τα οποία μπορούν να χρησιμοποιηθούν για αναλύσεις και ανάπτυξη υπηρεσιών.

Τα τηλεπικοινωνιακά δίκτυα παράγουν δεδομένα που σχετίζονται με τη χρήση δικτύων, την κινητικότητα των χρηστών, την ποιότητα υπηρεσιών και τη διαχείριση τηλεπικοινωνιακών υποδομών. Αυτά τα δεδομένα χρησιμοποιούνται συχνά για την ανάλυση κινητικότητας σε πόλεις ή για τη βελτιστοποίηση δικτύων επικοινωνίας.

Οι περιβαλλοντικοί αισθητήρες που εγκαθίστανται σε αστικά περιβάλλοντα ή φυσικά οικοσυστήματα συλλέγουν δεδομένα σε πραγματικό χρόνο για την ποιότητα του αέρα, τη θερμοκρασία, την υγρασία ή τα επίπεδα θορύβου. Τα δεδομένα αυτά αποτελούν σημαντική πηγή πληροφόρησης για την παρακολούθηση περιβαλλοντικών συνθηκών.

Παράλληλα, τα παραγόμενα δεδομένα που σχετίζονται με την κατανάλωση ενέργειας, την παραγωγή ηλεκτρικής ενέργειας και τη λειτουργία ενεργειακών δικτύων. Τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για την ανάπτυξη έξυπνων ενεργειακών συστημάτων και τη βελτίωση της ενεργειακής αποδοτικότητας.

2.3.3 Κατηγορίες Δεδομένων

Η συνεχής παραγωγή δεδομένων οδηγεί στη δημιουργία διαφορετικών κατηγοριών συνόλων δεδομένων, τα οποία μπορούν να χρησιμοποιηθούν για ανάλυση και ανάπτυξη εφαρμογών σε διάφορους επιστημονικούς τομείς, μπορούμε να διακρίνουμε την αξιοποίηση τους ανάλογα με το είδος στο οποίο ανήκουν.

- Τα περιβαλλοντικά δεδομένα περιλαμβάνουν πληροφορίες που σχετίζονται με το φυσικό περιβάλλον, όπως η ποιότητα του νερού, το αστικό πράσινο, πληροφορίες σχετικά με την πανίδα, χλωρίδα διαφόρων περιοχών. Τα ατμοσφαιρικά δεδομένα αφορούν την ποιότητα του αέρα, τις εκπομπές ρύπων και τις μετεωρολογικές συνθήκες, και χρησιμοποιούνται συχνά για την παρακολούθηση της ρύπανσης και των κλιματικών φαινομένων.
- Τα πληθυσμιακά δεδομένα παρέχουν πληροφορίες για τη δημογραφική σύνθεση των πληθυσμών, την κατανομή κατοίκων σε αστικές περιοχές και τις κοινωνικές δομές.
- Τα ενεργειακά δεδομένα περιλαμβάνουν στοιχεία σχετικά με την παραγωγή, κατανάλωση και διανομή ενέργειας, τα οποία είναι κρίσιμα για τον ενεργειακό σχεδιασμό και τη μετάβαση σε βιώσιμα ενεργειακά συστήματα.
- Τέλος, τα κλιματικά και τα δεδομένα κινητικότητας σχετίζονται με την παρακολούθηση του κλίματος και των μετακινήσεων σε αστικά περιβάλλοντα, παρέχοντας πολύτιμες πληροφορίες για τον αστικό σχεδιασμό, τις μεταφορές και την ανάπτυξη έξυπνων πόλεων.

Συνολικά, η συνεχής παραγωγή δεδομένων από αυτές τις διαφορετικές πηγές δημιουργεί ένα πλούσιο οικοσύστημα πληροφοριών, το οποίο μπορεί να αξιοποιηθεί για επιστημονική έρευνα, ανάπτυξη τεχνολογικών εφαρμογών και υποστήριξη τεκμηριωμένων πολιτικών αποφάσεων

2.4 Ο Σημασιολογικός Ιστός (Semantic Web)

Ο Σημασιολογικός Ιστός (Semantic Web) αποτελεί ένα από τα θεμελιώδη παραδείγματα οργάνωσης πληροφορίας στον Παγκόσμιο Ιστό, αντιπροσωπεύοντας την εξέλιξή του από ένα μέσο παρουσίασης εγγράφων σε ένα περιβάλλον διαμοιρασμού δεδομένων με νόημα. Στο πλαίσιο αυτό, ο στόχος υπερβαίνει την απλή μηχανικά αναγνώσιμη αποθήκευση πληροφοριών,

εστιάζοντας στη δημιουργία ενός σημασιολογικού πλαισίου που καθιστά το περιεχόμενο κατανοητό τόσο από τους ανθρώπους όσο και από τις μηχανές[14], [15].

Ο όρος εισήχθη το 2001 από τον Tim Berners-Lee, ο οποίος οραματίστηκε ένα δίκτυο δεδομένων διασυνδεδεμένων με τέτοιο τρόπο ώστε να επιτρέπεται η αυτόματη επεξεργασία, η αναζήτηση και η ανάλυσή τους από υπολογιστικά συστήματα [14]. Η βασική φιλοσοφία του Σημασιολογικού Ιστού έγκειται στην προσθήκη μεταδεδομένων και στη χρήση προτύπων που επιτρέπουν την περιγραφή σχέσεων μεταξύ των οντοτήτων, μετατρέποντας τον Παγκόσμιο Ιστό σε μια παγκόσμια, ενοποιημένη βάση γνώσης όπου η πληροφορία μπορεί να ανακτάται και να συνδυάζεται αυτόνομα από ευφυείς πράκτορες (agents)[15].

2.4.1 Resource Description Framework (RDF)

Το RDF (Resource Description Framework) αποτελεί το βασικό μοντέλο δεδομένων του Σημασιολογικού Ιστού, στο οποίο βασίζεται η ανάπτυξη δομημένης γνώσης[14]. Στο RDF, κάθε πληροφορία εκφράζεται με τη μορφή τριάδων (triples) της μορφής:

Υποκείμενο - Κατηγόρημα – Αντικείμενο-(Subject - Predicate - Object)

Κάθε στοιχείο αυτής της τριάδας αναγνωρίζεται μοναδικά από ένα URI (Uniform Resource Identifier), εξασφαλίζοντας ότι κάθε οντότητα είναι παγκοσμίως μοναδική και αναφέρσιμη. Για παράδειγμα, η πρόταση "Η Θεσσαλονίκη βρίσκεται στην Ελλάδα" μπορεί να αναπαρασταθεί ως:

<dbpedia:Thessaloniki> <dbpedia-owl:country> <dbpedia:Greece>

Η δομή RDF επιτρέπει τη δημιουργία γράφων (graphs) που συνδέουν οντότητες μεταξύ τους μέσω σχέσεων, δημιουργώντας πλούσιες δομές γνώσης. Το RDF υποστηρίζεται από πολλαπλές γλώσσες σειριοποίησης, όπως το Turtle (.ttl), το JSON-LD και το RDF/XML.

2.4.2 SPARQL - Η Γλώσσα Ερωτημάτων του RDF

Το SPARQL (SPARQL Protocol and RDF Query Language) αποτελεί τη γλώσσα ερωτημάτων για RDF δεδομένα, ανάλογη με την SQL για σχεσιακές βάσεις δεδομένων. Επιτρέπει την αναζήτηση και εξαγωγή πληροφοριών από RDF datasets και SPARQL endpoints. Η ύπαρξη SPARQL endpoints στο TDS και σε ευρωπαϊκές πλατφόρμες ανοικτών δεδομένων επιτρέπει την αυτόματη ανακάλυψη datasets μέσα από workflows.

2.4.3 OWL και RDFS - Γλώσσες Οντολογιών

Για την ανάπτυξη πλουσιότερων μοντέλων γνώσης, πάνω από το επίπεδο του RDF λειτουργούν οι γλώσσες έκφρασης οντολογιών. Το RDFS (RDF Schema) παρέχει τους βασικούς μηχανισμούς για την ιεραρχική περιγραφή κλάσεων και ιδιοτήτων. Η γλώσσα OWL (Web Ontology Language) επεκτείνει τις δυνατότητες του RDFS, προσφέροντας εκφραστικότερους μηχανισμούς που επιτρέπουν τον ορισμό σύνθετων σχέσεων, όπως η ισοδυναμία κλάσεων, οι μεταβατικές ιδιότητες και οι περιορισμοί στα πεδία τιμών. Αυτές οι τεχνολογίες αποτελούν τη βάση για την ανάπτυξη εξειδικευμένων οντολογιών, όπως η MDAT [10], η οποία χρησιμοποιείται για την περιγραφή και τη διαχείριση των δεδομένων στο πλαίσιο της παρούσας μελέτης.

2.5 Linked Data

Τα Συνδεδεμένα Δεδομένα (Linked Data) αποτελούν ένα σύνολο βέλτιστων πρακτικών για τη δημοσίευση και τη διασύνδεση δομημένων δεδομένων στον Παγκόσμιο Ιστό[14]. Η τεχνολογική τους βάση στηρίζεται στο μοντέλο RDF, ωστόσο η ουσιαστική τους αξία έγκειται

στη μετάβαση από έναν «Ιστό Εγγράφων» σε έναν «Ιστό Δεδομένων». Η φιλοσοφία αυτή κωδικοποιήθηκε από τον Tim Berners-Lee το 2006 μέσω τεσσάρων θεμελιωδών αρχών (Linked Data Principles), οι οποίες καθορίζουν τον τρόπο με τον οποίο οι πληροφορίες πρέπει να ταυτοποιούνται και να διαμοιράζονται:

1. Χρήση URIs: Κάθε οντότητα (πράγμα, έννοια ή τοποθεσία) πρέπει να προσδιορίζεται από ένα μοναδικό URI (Uniform Resource Identifier), λειτουργώντας ως μια παγκόσμια ταυτότητα.
2. Χρήση HTTP URIs: Τα αναγνωριστικά πρέπει να ακολουθούν το πρωτόκολλο HTTP, ώστε οι χρήστες και οι μηχανές να μπορούν να τα αναζητούν και να αποκτούν πρόσβαση σε αυτά.
3. Παροχή Χρήσιμης Πληροφορίας: Κατά την αναζήτηση ενός URI, το σύστημα πρέπει να επιστρέφει ουσιαστική πληροφορία για την οντότητα, χρησιμοποιώντας πρότυπα όπως το RDF και τη γλώσσα ερωτημάτων SPARQL.
4. Διασύνδεση: Κάθε περιγραφή δεδομένων πρέπει να περιλαμβάνει συνδέσμους προς άλλα URIs, επιτρέποντας την πλοήγηση και την ανακάλυψη νέας γνώσης μέσα από ένα παγκόσμιο δίκτυο σχέσεων.

Η συνδυαστική εφαρμογή αυτών των αρχών έχει οδηγήσει στη δημιουργία του Linked Open Data (LOD) Cloud, ενός τεράστιου, διασυνδεδεμένου δικτύου ανοικτών συνόλων δεδομένων που εκτείνεται σε παγκόσμια κλίμακα. Κεντρικοί κόμβοι αυτού του δικτύου είναι διεθνώς αναγνωρισμένα datasets, όπως η DBpedia (η δομημένη έκδοση της Wikipedia), το Wikidata και το GeoNames. Αυτές οι πηγές συνδέονται μεταξύ τους μέσω ειδικών σημασιολογικών ιδιοτήτων, όπως η owl:sameAs, η οποία δηλώνει ότι δύο διαφορετικά URIs αναφέρονται στην ίδια οντότητα του πραγματικού κόσμου.

Στο πλαίσιο της παρούσας εργασίας, οι αρχές των Linked Data αξιοποιούνται για τη σημασιολογική διασύνδεση των datasets του Thessaloniki Data Space (TDS) με την οντολογία MDAT. Με αυτόν τον τρόπο, τα δεδομένα παύουν να αποτελούν απομονωμένες νησίδες πληροφορίας (data silos) και εντάσσονται σε ένα ευρύτερο σημασιολογικό πλαίσιο. Η προσέγγιση αυτή επιτρέπει την αυτόματη ανακάλυψη, την ενοποίηση και τον συνδυασμό ετερογενών πηγών, διευκολύνοντας την παραγωγή σύνθετης γνώσης μέσα από το οικοσύστημα του UPCAST

2.6 Οντολογίες (Ontologies)

2.6.1 Τι είναι Οντολογία;

Στο πεδίο της Επιστήμης των Υπολογιστών, η οντολογία ορίζεται ως ένας επίσημος και τυπικός προσδιορισμός των εννοιών και των μεταξύ τους σχέσεων που διέπουν ένα συγκεκριμένο πεδίο γνώσης. Ενώ ο όρος έχει τις ρίζες του στη Φιλοσοφία, όπου μελετά τη φύση της ύπαρξης, στην πληροφορική αποκτά έναν αμιγώς πρακτικό και λειτουργικό ρόλο: τη δημιουργία ενός κοινού και μηχανικά αναγνώσιμου «λεξιλογίου». Μέσω των οντολογιών, η πληροφορία παύει να είναι ένα απλό σύνολο δεδομένων και αποκτά σημασιολογικό βάθος, επιτρέποντας στα υπολογιστικά συστήματα να «κατανοούν» το πλαίσιο στο οποίο αναφέρονται.[15]

Μια δομημένη οντολογία απαρτίζεται από τέσσερα βασικά συστατικά στοιχεία:

- Κλάσεις (Classes): Αποτελούν τις γενικές κατηγορίες ή τους τύπους των οντοτήτων στο υπό μελέτη πεδίο. Για παράδειγμα, στην οντολογία του UPCAST, οι κλάσεις

mdat:Dataset και mdat:ProcessingStep ορίζουν τις έννοιες του συνόλου δεδομένων και του βήματος επεξεργασίας αντίστοιχα.

- Ιδιότητες (Properties): Εκφράζουν τις σχέσεις μεταξύ των κλάσεων ή τα χαρακτηριστικά τους. Χαρακτηριστικά παραδείγματα αποτελούν οι ιδιότητες mdat:hasInput και mdat:producesOutput, οι οποίες συνδέουν τις διαδικασίες με τα δεδομένα που καταναλώνουν ή παράγουν.
- Άτομα (Individuals): Πρόκειται για τα συγκεκριμένα στιγμιότυπα (instances) των κλάσεων που αναπαριστούν οντότητες του πραγματικού κόσμου, όπως το σύνολο δεδομένων mdat:UrbanTreeCategoriesDataset.
- Αξιώματα (Axioms): Αποτελούν τους λογικούς κανόνες και τους περιορισμούς που διασφαλίζουν τη συνοχή της γνώσης. Ένα αξίωμα μπορεί να επιβάλλει, για παράδειγμα, ότι κάθε βήμα επεξεργασίας (ProcessingStep) πρέπει υποχρεωτικά να διαθέτει τουλάχιστον μία είσοδο δεδομένων, επιτρέποντας έτσι στο σύστημα να εντοπίζει λογικά σφάλματα στον σχεδιασμό μιας ροής εργασιών

2.6.2 Γιατί Χρειαζόμαστε Οντολογίες στα Data Spaces;

Η αναγκαιότητα των οντολογιών γίνεται ιδιαίτερα εμφανής στα Οικοσυστήματα Δεδομένων (Data Spaces), όπου συμμετέχει πλήθος διαφορετικών παρόχων με ετερογενή σχήματα δεδομένων. Η υιοθέτηση μιας κοινής οντολογίας αποτελεί τον καταλύτη για την εύρυθμη λειτουργία τέτοιων περιβαλλόντων, προσφέροντας τα εξής πλεονεκτήματα: Πρώτον, εξασφαλίζεται ενιαία σημασιολογία, διασφαλίζοντας ότι όλοι οι συμμετέχοντες ερμηνεύουν τους όρους (π.χ. «πληθυσμός» ή «ρύπος») με τον ίδιο ακριβώς τρόπο, εκμηδενίζοντας τις ασάφειες. Δεύτερον, διευκολύνεται η αυτόματη ανακάλυψη δεδομένων, καθώς οι αλγόριθμοι αναζήτησης μπορούν να εντοπίζουν σχετικά datasets βάσει της σημασίας τους και όχι μόνο βάσει λέξεων-κλειδιών. Επιπλέον, επιτυγχάνεται η σημασιολογική διαλειτουργικότητα, η οποία επιτρέπει τη δυναμική σύνθεση ροών εργασίας (workflows), όπου η έξοδος μιας διαδικασίας αναγνωρίζεται αυτόματα ως έγκυρη είσοδος για την επόμενη. Τέλος, η χρήση οντολογιών ενισχύει την αναλυσιμότητα του συστήματος. Μέσω μηχανών συλλογιστικής (reasoning engines), είναι εφικτή η εξαγωγή νέας, έμμεσης γνώσης από τις υπάρχουσες σχέσεις, επιτρέποντας στο Data Space να «αντιλαμβάνεται» σύνθετες αλληλεξαρτήσεις που δεν είναι άμεσα ορατές στα πρωτογενή δεδομένα[8].

2.6.3 Η Οντολογία MDAT

Στο πλαίσιο της παρούσας εργασίας αναπτύχθηκε η οντολογία MDAT (με το αντίστοιχο namespace mdat:), η οποία αποτελεί το βασικό εννοιολογικό μοντέλο για την περιγραφή των πόρων του συστήματος. Η MDAT παρέχει έναν δομημένο και επίσημο ορισμό για ένα ευρύ φάσμα οντοτήτων, συμπεριλαμβανομένων των όρων δεδομένων, των συνόλων δεδομένων (datasets), των λειτουργιών επεξεργασίας, των οργανωτικών ρόλων, καθώς και των παράγωγων δεικτών.

Ενδεικτικά, η οντολογία περιλαμβάνει εξειδικευμένες έννοιες όπως:

- mdat:Population Distribution: Αναπαριστά το σύνολο δεδομένων της πληθυσμιακής κατανομής ανά δημοτική ενότητα.
- mdat:TreesDistribution: Ορίζει το σύνολο δεδομένων που αφορά τη γεωγραφική κατανομή των αστικών δέντρων.
- mdat:ExtractPopulationCount: Περιγράφει τη συγκεκριμένη λειτουργία επεξεργασίας για την εξαγωγή πληθυσμιακών αριθμών.

- **mdat:UrbanGreenIndicators:** Προσδιορίζει τον παράγωγο δείκτη που υπολογίζει την αναλογία αστικού πρασίνου ανά κάτοικο.

Η οντολογία MDAT λειτουργεί ως η «σημασιολογική ραχοκοκαλιά» του προτεινόμενου συστήματος, καθώς πάνω σε αυτή βασίζεται η μοντελοποίηση των ροών εργασίας (workflows), η διαπραγμάτευση πολιτικών πρόσβασης και οι μηχανισμοί αυτόματης επανασχεδίασης των διεργασιών. Η πλήρης τεχνική περιγραφή και ο ορισμός της οντολογίας είναι δημόσια διαθέσιμα στο αποθετήριο: <https://github.com/MariannaVouraki/Mdat-Ontology-definement> [3].

2.7 Data Privacy Vocabulary (DPV)

2.7.1 Τι είναι το DPV;

Το Λεξιλόγιο Ιδιωτικότητας Δεδομένων (Data Privacy Vocabulary - DPV) [7] αποτελεί μια τεχνική προδιαγραφή που αναπτύχθηκε από την κοινότητα W3C Data Privacy Vocabularies and Controls Community Group (DPVCG). Πρωταρχικός στόχος του DPV είναι η παροχή ενός κοινού, τυποποιημένου και μηχανικά αναγνώσιμου λεξιλογίου για την έκφραση εννοιών που σχετίζονται με την επεξεργασία προσωπικών δεδομένων. Η δομή του είναι πλήρως ευθυγραμμισμένη με τις απαιτήσεις του Γενικού Κανονισμού Προστασίας Δεδομένων (GDPR - Κανονισμός (ΕΕ) 2016/679) [6], διασφαλίζοντας τη νομική και τεχνική συνέπεια κατά τη διαχείριση της πληροφορίας [7].

- Το DPV οργανώνει τη γνώση γύρω από την ιδιωτικότητα ορίζοντας κεντρικές έννοιες που αφορούν:
- Κατηγορίες Δεδομένων: Παρέχει διακρίσεις μεταξύ διαφορετικών τύπων δεδομένων, όπως προσωπικά (*dpv:PersonalData*), ευαίσθητα (*dpv:SensitiveData*) και ανωνυμοποιημένα δεδομένα (*dpv:AnonymisedData*).
- Σκοπούς Επεξεργασίας: Προσδιορίζει το πλαίσιο χρήσης της πληροφορίας, περιλαμβάνοντας έννοιες όπως η έρευνα και ανάπτυξη (*dpv:ResearchAndDevelopment*) και η μη εμπορική έρευνα (*dpv:NonCommercialResearch*).
- Νομικές Βάσεις: Κωδικοποιεί τις προϋποθέσεις νομιμότητας της επεξεργασίας, όπως η συγκατάθεση (*dpv:Consent*), η νομική υποχρέωση (*dpv:LegalObligation*) και το έννομο συμφέρον (*dpv:LegitimateInterest*).
- Λειτουργίες Επεξεργασίας: Περιγράφει λεπτομερώς τις τεχνικές ενέργειες επί των δεδομένων, από τη συλλογή (*dpv:Collect*) και την αποθήκευση (*dpv:Store*), έως τη διαγραφή (*dpv:Erase*), τον μετασχηματισμό (*dpv:Transform*), την ομαδοποίηση (*dpv:Aggregate*), την παραγωγή νέας γνώσης (*dpv:Derive*) και την οπτικοποίηση (*dpv:Visualise*).
- Οργανωτικούς Ρόλους: Ταυτοποιεί τα εμπλεκόμενα μέρη, ορίζοντας τον Υπεύθυνο Επεξεργασίας (*dpv:DataController*), τον Εκτελούντα την Επεξεργασία (*dpv:DataProcessor*) και το Υποκείμενο των Δεδομένων (*dpv:DataSubject*).

Η υιοθέτηση του DPV επιτρέπει τη σημασιολογική μοντελοποίηση των πολιτικών προστασίας, καθιστώντας τον έλεγχο της συμμόρφωσης μια αυτοματοποιημένη και διαφανή διαδικασία εντός των ψηφιακών οικοσυστημάτων δεδομένων.

2.7.2 DPV στα Workflows Επεξεργασίας Δεδομένων

Ένα από τα πιο σημαντικά πλεονεκτήματα του DPV είναι η ικανότητά του να εκφράζει νομικές υποχρεώσεις ως εκτελέσιμες λειτουργίες. Στο πλαίσιο της παρούσας εργασίας, κάθε βήμα του DPW αντιστοιχίζεται σε μία ή περισσότερες `dpv:Processing operations`:

- Ένα βήμα καθαρισμού δεδομένων → `dpv:Transform`
- Ένα βήμα υπολογισμού δεικτών → `dpv:Derive`, `dpv:Aggregate`
- Ένα βήμα οπτικοποίησης → `dpv:Visualise`, `dpv:Analyse`
- Ένα αυτόματα εισαχθέν βήμα διαγραφής → `dpv:Erase`

Αυτή η αντιστοίχιση επιτρέπει στο σύστημα να γνωρίζει ανά πάσα στιγμή ποιες επεξεργασίες πραγματοποιούνται και να επαληθεύει αν αυτές είναι σύμφωνες με τις δηλωμένες πολιτικές.

2.8 Open Digital Rights Language (ODRL)

2.8.1 Τι είναι το ODRL;

Το Open Digital Rights Language (ODRL) είναι ένα πρότυπο του W3C για την έκφραση πολιτικών χρήσης (*usage policies*) σε μηχανικώς αναγνώσιμη, RDF-based μορφή [9], [16]. Επιτρέπει σε παρόχους δεδομένων να ορίζουν επακριβώς ποιος μπορεί να κάνει τι με τα δεδομένα τους, υπό ποιες συνθήκες και με ποιες υποχρεώσεις.

Ένα ODRL policy αποτελείται από τρία βασικά στοιχεία:

- **Permission (Άδεια):** Τι επιτρέπεται (π.χ. `odrl:use`, `odrl:reproduce`, `odrl:derive`)
- **Prohibition (Απαγόρευση):** Τι απαγορεύεται (π.χ. `odrl:commercialUse`)
- **Obligation (Υποχρέωση):** Τι πρέπει να γίνει (π.χ. `odrl:delete` μετά τη χρήση, `odrl:attribute` για αναφορά πηγής)

2.8.2 ODRL Policy Παράδειγμα

Ένα τυπικό ODRL policy για ένα dataset με ODbL άδεια μπορεί να εκφράζεται:

"Ο χρήστης μπορεί να χρησιμοποιεί και να παράγει παράγωγα (`odrl:use`, `odrl:derive`), υπό την υποχρέωση αναφοράς της πηγής (`odrl:attribute`) και χρήσης ισοδύναμης άδειας στα παράγωγα."

Στο Σενάριο 2 αυτής της εργασίας, το ODRL policy του extended demographic dataset εκφράζει μια πολιτική μη-αποθήκευσης:

`odrl:Prohibition` → `odrl:store` (απαγορεύεται η αποθήκευση)

`odrl:Obligation` → `dpv:Erase` (υποχρεωτική διαγραφή μετά τη χρήση)

Το Consumer Dashboard ερμηνεύει αυτήν την πολιτική και εισάγει αυτόματα τον κόμβο `dpv:delete` στο workflow, υλοποιώντας τη λογική *Compliance-by-Design*.

2.8.3 Συνδυασμός ODRL + DPV για Compliance-by-Design

Η ισχύς της προσέγγισης αυτής της εργασίας έγκειται στον συνδυασμό ODRL και DPV: το ODRL ορίζει τι επιτρέπεται (*permissions/prohibitions/obligations*)[9], ενώ το DPV περιγράφει ποιες λειτουργίες εκτελούνται πραγματικά (*processing operations*)[7]. Το πλαίσιο DPW επαληθεύει αυτόματα ότι κάθε `dpv:Processing operation` είναι καλυμμένη από ένα αντίστοιχο `odrl:Permission`, και εισάγει αυτόματα τους υποχρεωτικούς κόμβους *compliance*[5], [17]. Αυτή είναι η ουσία του *Compliance-by-Design* αφορά την συμμόρφωση του συστήματος σε

πολιτικές που δεν εφαρμόζονται από ανθρώπινη παρέμβαση, αλλά είναι ενσωματωμένες στην αρχιτεκτονική του συστήματος .

2.9 Data Processing Workflows: Έννοια και Αξία

2.9.1 Τι είναι Data Processing Workflow;

Ένα Data Processing Workflow (DPW) είναι μια δομημένη, μηχανικώς εκτελέσιμη ακολουθία βημάτων επεξεργασίας δεδομένων, στην οποία κάθε βήμα καθορίζεται με ακρίβεια ως προς τις εισόδους, τις εξόδους, τις λειτουργίες, τους ρόλους και τις σημασιολογικές συνδέσεις. Αντιθέτως από ένα απλό script, ένα DPW:

- Ορίζει ρητά την εξάρτηση μεταξύ βημάτων (DAG - Directed Acyclic Graph)
- Παρέχει σημασιολογική τεκμηρίωση για κάθε λειτουργία
- Επιτρέπει παράλληλη εκτέλεση ανεξάρτητων βημάτων
- Εξασφαλίζει αναπαραγωγιμότητα: το ίδιο workflow με τα ίδια δεδομένα παράγει πάντα τα ίδια αποτελέσματα
- Υποστηρίζει παρακολούθηση (monitoring) σε πραγματικό χρόνο

2.9.2 Η Αξία της Επαναχρησιμοποίησης Workflows

Ένα από τα κρίσιμα πλεονεκτήματα των workflows είναι η επαναχρησιμοποίησή τους. Όταν ένα workflow ολοκληρωθεί και επικυρωθεί, αποθηκεύεται ως πρότυπο (template) και μπορεί να:

- Επαναχρησιμοποιηθεί αμέσως για ένα νέο dataset: π.χ. η απογραφή του 2021 μπορεί να τρέξει αυτόματα μέσα από το ίδιο workflow Urban Green Indicators
- Παραμετροποιηθεί: Αλλαγή μόνο της πηγής δεδομένων, χωρίς επανασχεδιασμό της λογικής
- Κοινοποιηθεί: Άλλοι ερευνητές ή δήμοι μπορούν να χρησιμοποιήσουν το ίδιο workflow για τη δική τους πόλη
- Ελεγχθεί (audited): Κάθε εκτέλεση τεκμηριώνεται αυτόματα για λόγους διαφάνειας και ευθύνης

Η Fafalios et al. [18] τυποποιούν αυτή την προσέγγιση μέσω ενός μοντέλου workflow με έμφαση στον σημασιολογικό εμπλουτισμό. Η μηχανή Nextflow [19] υλοποιεί αυτή την αλληλουχία σε κλίμακα, υποστηρίζοντας αναπαραγωγίμες αναλυτικές pipelines. Η παρούσα εργασία ενσωματώνει και τις δύο αυτές επιστημονικές προσεγγίσεις.

Σχετική Βιβλιογραφία

Ολοκληρωμένα πλαίσια που ενοποιούν ενορχήστρωση workflow, σημασιολογική διαλειτουργικότητα και policy-aware εκτέλεση σε ένα ενιαίο περιβάλλον παραμένουν περιορισμένα [18]. Οι Kovach et al. [12] σε μια συστηματική ανασκόπηση της βιβλιογραφίας επιβεβαιώνουν ότι η έρευνα για τα Data Spaces παραμένει κατακερματισμένη μεταξύ διαλειτουργικότητας, διακυβέρνησης και ενορχήστρωσης. Οι Cimmino et al. [16] εισάγουν ένα πλαίσιο μετατροπής περιγραφικών πολιτικών σε επιβλητούς κανόνες - ένα χάσμα που η παρούσα εργασία καλύπτει ενοποιώντας και τα τρία στοιχεία σε ένα compliance-aware περιβάλλον εκτέλεσης[20], [21].

Κεφάλαιο 3ο: Το Οικοσύστημα UPCAST

3.1 Το Thessaloniki Data Space (TDS)

Το Thessaloniki Data Space (TDS) είναι μια πλατφόρμα αποθήκευσης δεδομένων που ιδρύθηκε στο πλαίσιο του Ευρωπαϊκού Έργου UPCAST. Αναπτύχθηκε από το Open Knowledge Foundation Greece (OKFN Greece) και δημιουργήθηκε για να λειτουργήσει ως πλατφόρμα διαμεσολάβησης μεταξύ καταναλωτών (data consumers) και παρόχων δεδομένων (data providers). Φιλοξενεί κυρίως ελεύθερα διαθέσιμα σύνολα δεδομένων που προέρχονται από τη δημόσια διοίκηση, ανοικτές πηγές, ερευνητικά έργα και οργανισμούς που παράγουν δεδομένα στην πόλη της Θεσσαλονίκης.

Τα περισσότερα από τα διαθέσιμα σύνολα δεδομένων αφορούν περιβαλλοντικές πτυχές. Σύμφωνα με την κατάσταση του TDS κατά τη συγγραφή της παρούσας εργασίας (Ιανουάριος 2026), υπάρχουν διαθέσιμα 123 περιβαλλοντικά datasets, 50 datasets που αφορούν την Υγεία, 30 για την επιστήμη και την τεχνολογία, καθώς και άλλες κατηγορίες. Οι διαθέσιμες κατηγορίες περιλαμβάνουν το νερό, την ποιότητα του αέρα, την ατμοσφαιρική ρύπανση, καταγραφές αποβλήτων, Urban Green Data και άλλα.

Στο παρασκήνιο βρίσκεται μια πληθώρα προηγμένων τεχνικών υποδομών (Technical Infrastructure). Το TDS ενσωμάτωσε εργαλεία για το policy negotiation, επιτρέποντας σε παρόχους και καταναλωτές να συμφωνούν σε συγκεκριμένους όρους χρήσης, licensing και δικαιώματα πρόσβασης (access rights). Διαθέτει ενισχυμένες δυνατότητες αναζήτησης μέσω του discovery plugin και παρέχει εκτίμηση περιβαλλοντικού αποτυπώματος (environmental impact estimation) για κάθε dataset μέσω του environmental plugin. Η πλατφόρμα υποστηρίζει επίσης τη λειτουργία Dataset on Demand, δίνοντας στους χρήστες τη δυνατότητα να ζητήσουν ένα συγκεκριμένο σύνολο δεδομένων εάν απαιτείται· αφού ο καταναλωτής υποβάλει ένα αίτημα, η πλατφόρμα το προωθεί σε όλους τους πιθανούς παρόχους.

3.2 Τα Εργαλεία UPCAST

Το TDS που αναφέραμε χρησιμοποιεί εργαλεία που δημιουργήσαμε στο πλαίσιο του έργου UPCAST (Universal Platform Components for Safe Fair Interoperable Data Exchange, Monetization and Trading) [13] [22]. Αυτό το έργο ανέπτυξε μια σουίτα από modular plugins σχεδιασμένα να αυτοματοποιούν και να θωρακίζουν την κυκλική διαδικασία ανταλλαγής δεδομένων. Τα εργαλεία αυτά αναπτύχθηκαν για να ενσωματωθούν (integrated) σε υπάρχουσες αγορές δεδομένων, όπως το Thessaloniki Data Space. Στόχος τους είναι η αυτοματοποίηση της χειροκίνητης διαδικασίας σύναψης συμβολαίων για την αγορά, την πώληση και κυρίως την απόκτηση συνόλων δεδομένων. Τα κύρια plug-ins που δημιουργήθηκαν περιλαμβάνουν:

- **Discovery Plugin:** Επιτρέπει την προηγμένη έρευνα δεδομένων χρησιμοποιώντας σημασιολογικά φίλτρα (semantic search), παρέχοντας στον χρήστη ακριβέστερα αποτελέσματα.
- **Pricing Plugin:** Χρησιμοποιεί αλγορίθμους τεχνητής νοημοσύνης (AI algorithms) για να προτείνει δίκαιες τιμές για την πώληση ή τη χρήση δεδομένων, καθώς και την αντιληπτή αξία τους με βάση την ποιότητα και τη ζήτηση.
- **Negotiation Plugin:** Αυτοματοποιεί τη διαδικασία συμφωνίας μεταξύ παρόχου και καταναλωτή σχετικά με τους όρους χρήσης και το κόστος των δεδομένων μέσω ψηφιακών συμβολαίων.

- **Policy Editor Plugin:** Επιτρέπει στους παρόχους να ορίζουν εύκολα κανόνες προστασίας (π.χ. ποιος μπορεί να δει τα δεδομένα και για πόσο χρονικό διάστημα) σε μορφή που είναι «αναγνώσιμη» από υπολογιστή.
- **Consent Manager Plugin:** Διαχειρίζεται τις άδειες που δίνονται από τους πολίτες για τη χρήση των προσωπικών τους δεδομένων, διασφαλίζοντας τη συμμόρφωση με τον GDPR.
- **Environmental Impact Plugin:** Υπολογίζει το μέγεθος του περιβαλλοντικού αποτυπώματος και τις εκπομπές CO₂ που προκύπτουν κατά την αποθήκευση και τη μεταφορά συγκεκριμένων συνόλων δεδομένων (datasets).
- **Federated Machine Learning Plugin:** Επιτρέπει σε μοντέλα AI να μαθαίνουν από τα δεδομένα χωρίς αυτά να γίνονται ορατά σε τρίτους, διατηρώντας έτσι τα ευαίσθητα δεδομένα σας ασφαλή.
- **Data Integration & Exchange Plugin:** Διασφαλίζει τη διαλειτουργικότητα (interoperability) μεταξύ διαφορετικών πηγών και μορφοτύπων δεδομένων, εξασφαλίζοντας ότι μπορούν να επικοινωνούν και να μεταφέρονται με ασφάλεια.

Συνολικά, το UPCAST δεν αφορά μόνο την παροχή τεχνικών εργαλείων. Οικοδομεί έναν εντελώς νέο τρόπο λειτουργίας των ψηφιακών οικοσυστημάτων, διασφαλίζει ότι η πληροφορία ανταλλάσσεται με ασφάλεια, μπορεί να χρησιμοποιηθεί σε διαφορετικά συστήματα, λαμβάνει υπόψη περιβαλλοντικούς παράγοντες και συμμορφώνεται με τους κανόνες του GDPR. Σημειώνεται ότι, στο πλαίσιο της παρούσας εργασίας, αξιοποιούνται κατά κύριο λόγο τα Discovery, Negotiation, Policy Editor και Environmental Impact Plugins, ενώ τα υπόλοιπα (Pricing, Consent Manager, Federated Machine Learning, Data Integration & Exchange) παρουσιάζονται για λόγους πληρότητας του οικοσυστήματος.

3.3 Consumer Dashboard

Το Consumer Dashboard λειτουργεί ως ο κεντρικός κόμβος ενορχήστρωσης (orchestration hub) του πλαισίου UPCAST. Αναπτύχθηκε αξιοποιώντας και επεκτείνοντας το ερευνητικό πρωτότυπο goodFlows της ICT Aboveo ([Solutions – goodFlows | ict aboveo](#)), και παρέχει στους καταναλωτές δεδομένων ένα ενιαίο γραφικό περιβάλλον εργασίας (GUI) για την αλληλεπίδραση με το σύνθετο οικοσύστημα των plugins και των backend services του UPCAST. Αντί για την αλληλεπίδραση με μεμονωμένα APIs, το dashboard κεντριοκοποιεί τη διαδρομή του καταναλωτή από την αρχική ανακάλυψη δεδομένων έως την εκτελέσιμη ανάλυση μέσα σε ένα ενιαίο περιβάλλον που βασίζεται στη ροή εργασιών (workflow-driven)[22].

Core Functionalities

Η πλατφόρμα ενσωματώνει πέντε βασικούς επιχειρησιακούς πυλώνες:

- **Semantic Modeling & Ontology Alignment:** Χρησιμοποιώντας έναν εξειδικευμένο μηχανισμό για την εισαγωγή και την ευθυγράμμιση με τις οντολογίες DPV και ODRL, το dashboard επιτρέπει στους χρήστες να μοντελοποιούν οργανωτικές πτυχές (ρόλους, σκοπούς και τύπους δεδομένων) ως σημασιογραφικά γραφήματα (semantic graphs). Αυτό διασφαλίζει ότι κάθε στοιχείο στο workflow χαρακτηρίζεται από σαφώς καθορισμένες ιεραρχίες και σχέσεις.
- **Policy Administration Point (PAP):** Το dashboard λειτουργεί ως ένα ολοκληρωμένο PAP, επιτρέποντας τη διαχείριση των εσωτερικών οργανωτικών πολιτικών και των κανονιστικών υποχρεώσεων. Αυτό διασφαλίζει ότι όλες οι πρακτικές δεδομένων

συμμορφώνονται τόσο με τους θεσμικούς κανόνες όσο και με νομικά πλαίσια όπως το GDPR.

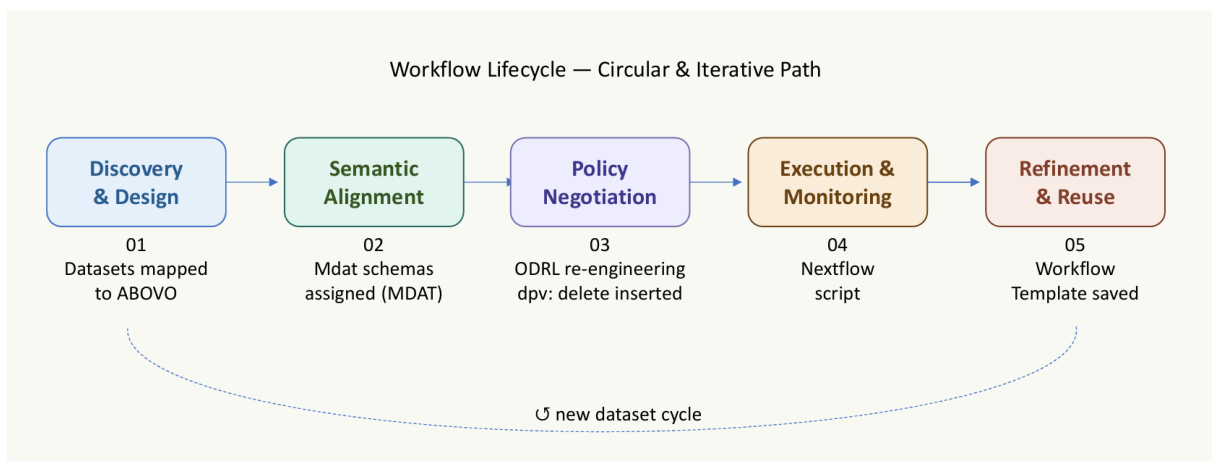
- **Visual Workflow Design:** Στην καρδιά του dashboard βρίσκεται ένα περιβάλλον μοντελοποίησης βασισμένο σε κόμβους (node-based modeling). Οι καταναλωτές μπορούν να σχεδιάσουν οπτικά τα Data Processing Workflows (DPW), καθορίζοντας παραμέτρους για τις εργασίες (actors, operations, assets) και τις ακμές (data flow and constraints) για τη διαχείριση των συνόλων δεδομένων καθ' όλη τη διάρκεια του κύκλου ζωής τους.
- **Verification and Workflow Re-engineering:** Ένα κρίσιμο χαρακτηριστικό είναι η αυτοματοποιημένη επαλήθευση (verification) των workflows έναντι τρίτων και εσωτερικών πολιτικών. Το dashboard εντοπίζει σημασιολογικές συγκρούσεις (semantic conflicts) και διευκολύνει την ανασχεδίαση της ροής εργασιών (workflow re-engineering), όπως η εισαγωγή υποχρεωτικών βημάτων διαγραφής ή η έναρξη διαπραγματεύσεων, ώστε να διασφαλιστεί η πλήρης συμμόρφωση πριν από την εκτέλεση.
- **NextFlow Execution Integration:** Για να γεφυρωθεί το χάσμα μεταξύ σχεδιασμού και δράσης, η πλατφόρμα μεταφράζει τα οντολογικά μοντέλα workflow σε εκτελέσιμα NextFlow scripts. Αυτή η backend ενσωμάτωση περιλαμβάνει εντολές ειδικά για το UPCAST για την παρακολούθηση της εκτέλεσης σε πραγματικό χρόνο (real-time execution monitoring) και την παρακολούθηση της απόδοσης.

Διασυνδέοντας απευθείας με τα plugins των Discovery, Negotiation, Policy Editor και Environmental Impact, το Consumer Dashboard μετατρέπει τις αφηρημένες πολιτικές δεδομένων σε δομημένους, επαναχρησιμοποιήσιμους και συμμορφούμενους αγωγούς επεξεργασίας δεδομένων (data processing pipelines).

3.4 Το Πλαίσιο Data Processing Workflow (DPW) εντός του UPCAST

3.4.1 Αρχιτεκτονική και Κύκλος Ζωής DPW

Ο κύκλος ζωής ενός DPW εντός του UPCAST ecosystem ακολουθεί μια δομημένη, κυκλική και επαναλαμβανόμενη διαδικασία πέντε σταδίων, που εξασφαλίζει ότι η ανάλυση δεν είναι απλώς εφάπαξ γεγονός αλλά αναπαραγωγική υπηρεσία.



Σχήμα 3.1: Κύκλος Ζωής DPW - Κυκλική και Επαναληπτική Διαδρομή

Στάδιο 1: Discovery and Design

Στο πρώτο στάδιο, ο καταναλωτής δεδομένων (Data Consumer) εκκινεί τη διαδικασία εντοπίζοντας τα κατάλληλα σύνολα δεδομένων που απαιτούνται για την αναλυτική του ανάγκη. Η αναζήτηση γίνεται μέσω του Discovery Plugin του UPCAST, το οποίο αξιοποιεί σημασιολογικά φίλτρα ερωτήματα επί του Thessaloniki Data Space. Με τον τρόπο αυτό, τα datasets εντοπίζονται βάσει της σημασιολογικής τους κατηγοριοποίησης (π.χ. `dpv:EnvironmentalData`, `mdat:PopulationDistribution`) και όχι μόνο βάσει τίτλου ή λέξεων-κλειδιών.

Τα επιλεγμένα datasets χαρτογραφούνται στο Consumer Dashboard ως αρχικοί κόμβοι εισόδου του υπό σχεδιασμό workflow, ενώ ο καταναλωτής καθορίζει την τοπολογία της ροής, ορίζοντας τα ενδιάμεσα βήματα επεξεργασίας και τις μεταξύ τους εξαρτήσεις δεδομένων. Το αποτέλεσμα του σταδίου είναι ένα οντολογικό μοντέλο workflow, αποτυπωμένο σε JSON-LD, το οποίο περιγράφει τη ροή.

Στάδιο 2: Semantic Alignment

Στο δεύτερο στάδιο πραγματοποιείται η σημασιολογική ευθυγράμμιση των βημάτων του workflow με τις οντολογίες αναφοράς. Κάθε κόμβος αντιστοιχίζεται σε μία ή περισσότερες κλάσεις του DPV (π.χ. `dpv:Transform`, `dpv:Aggregate`, `dpv:Derive`), σε αντίστοιχα actions του ODRL (π.χ. `odrl:modify`, `odrl:aggregate`, `odrl:derive`) και σε εξειδικευμένες έννοιες του namespace `mdat`: για όσα στοιχεία απαιτούν εξειδίκευση πεδίου που υπερβαίνει τους γενικούς όρους.

Η σημασιολογική αυτή αντιστοίχιση εξασφαλίζει τρία πράγματα: πρώτον, ότι οι έξοδοι ενός βήματος είναι συμβατές με τις εισόδους του επόμενου (semantic interoperability) π.χ. η έξοδος `mdat:PopulationCount` από το `ExtractPopulationCount` αναγνωρίζεται αυτόματα ως έγκυρη είσοδος του `CalculateUrbanGreenIndicators`. Δεύτερον, ότι η ροή είναι αναγνωρίσιμη και επεξεργάσιμη από μηχανές συλλογιστικής (reasoners), επιτρέποντας τον αυτόματο εντοπισμό σημασιολογικών συγκρούσεων. Τρίτον, ότι το workflow γίνεται αναζητήσιμο και επαναχρησιμοποιήσιμο ως πρότυπο επί της οντολογικής του αναπαράστασης.

Στάδιο 3: Policy Negotiation and Re-engineering

Το τρίτο στάδιο αποτελεί τον πυρήνα της καινοτομίας του πλαισίου DPW. Πριν από οποιαδήποτε εκτέλεση, το consumer Dashboard ανακτά τα ODRL policies που συνοδεύουν κάθε επιλεγμένο dataset και τα αναλύει συστηματικά μέσω ενός μηχανισμού επαλήθευσης. Η διαδικασία περιλαμβάνει τρεις διακριτές υπο-φάσεις:

- **Policy Retrieval and Parsing.** Το σύστημα ανακτά τα ODRL policies από το SPARQL endpoint του TDS ή από επισυναπτόμενα μεταδεδομένα του dataset. Τα policies αναλύονται σε τρεις βασικές κατηγορίες οντοτήτων: `odrl:Permission` (επιτρεπτές ενέργειες), `odrl:Prohibition` (απαγορεύσεις) και `odrl:Obligation` (υποχρεώσεις που πρέπει να εκπληρωθούν για να ισχύσει η άδεια χρήσης).
- **Compliance Verification.** Για κάθε βήμα του workflow, το σύστημα ελέγχει αν η αντίστοιχη `dpv:Processing operation` καλύπτεται από μία τουλάχιστον `odrl:Permission` του υποκείμενου dataset. Εάν εντοπιστεί ασυμφωνία π.χ. βήμα που εκτελεί `dpv:Derive` σε dataset του οποίου η πολιτική επιτρέπει μόνο `odrl:use` εκκινεί διαδικασία διαπραγμάτευσης μέσω του UPCAST Negotiation Plugin (όπως αναλυτικά παρουσιάζεται στο Σενάριο 2 του Κεφαλαίου 5).
- **Automatic Workflow Re-engineering.** Εάν στα policies εντοπιστούν υποχρεώσεις, το σύστημα εισάγει αυτόματα νέους κόμβους στο workflow που εξασφαλίζουν τη

συμμόρφωση. Τυπικές τέτοιες ενέργειες είναι η εισαγωγή κόμβου `dpv:Erase` (με ODRL action `odrl:delete`) αμέσως μετά τη χρήση ευαίσθητων ενδιάμεσων δεδομένων, ή κόμβου `dpv:Store` (με `odrl:reproduce`) στο τέλος της ροής για τη διατήρηση μόνο των τελικών συγκεντρωτικών αποτελεσμάτων. Σημαντικό είναι ότι αυτή η επανασχεδίαση δεν απαιτεί παρέμβαση του χρήστη: το πλαίσιο αναγνωρίζει αυτόνομα τις υποχρεώσεις και τις μεταφράζει σε εκτελέσιμες ενέργειες.

Στάδιο 4: Execution and Monitoring

Το ευθυγραμμισμένο και επανασχεδιασμένο workflow μεταφράζεται σε εκτελέσιμο Nextflow script μέσω του NextFlow Execution Integration του ABOVO. Η επιλογή της Nextflow ως εκτελεστικής μηχανής δεν είναι τυχαία: η συμβατότητά της με containers εξασφαλίζει πλήρη αναπαραγωγικότητα του υπολογιστικού περιβάλλοντος, ενώ η εγγενής υποστήριξη παραλληλισμού επιτρέπει την ταυτόχρονη εκτέλεση ανεξάρτητων βημάτων (π.χ. στα Σενάρια 1 και 2, η εξυγίανση των πληθυσμιακών και περιβαλλοντικών datasets γίνεται παράλληλα).

Κατά την εκτέλεση, το σύστημα παρακολουθεί τρεις βασικούς δείκτες: (α) την πρόοδο εκτέλεσης (status κάθε κόμβου), (β) τα σφάλματα και τις εξαιρέσεις μέσω του fault-tolerance μηχανισμού της Nextflow, και (γ) το υπολογιστικό αποτύπωμα άνθρακα της εκτέλεσης μέσω του Environmental Impact Plugin, ενός χαρακτηριστικού που εντάσσεται στη φιλοσοφία βιωσιμότητας του UPCAST. Όλες οι ενέργειες καταγράφονται σε audit log σε μορφή JSON-LD, εξασφαλίζοντας πλήρη ιχνηλασιμότητα.

Στάδιο 5: Refinement and Reuse

Στο τελικό στάδιο, τα παραγόμενα αποτελέσματα αξιολογούνται από τον Researcher ή τον Data Analyst ως προς την ποιότητα και την καταλληλότητά τους για τον επιδιωκόμενο σκοπό. Εάν απαιτηθούν προσαρμογές (π.χ. αλλαγή φίλτρων, προσθήκη επιπλέον βήματος επεξεργασίας), η ροή επανεισέρχεται στο Στάδιο 1.

Εάν το workflow κρίνεται ολοκληρωμένο, αποθηκεύεται ως πρότυπο (template) στη βιβλιοθήκη του Consumer Dashboard, διατηρώντας πλήρη τη σημασιολογική του τεκμηρίωση και τα ODRL policies. Η αξία της προσέγγισης αυτής αναδεικνύεται όταν εμφανιστεί νέο dataset της ίδιας κατηγορίας - για παράδειγμα, η Απογραφή Πληθυσμού 2031 που θα διαδεχθεί την Απογραφή 2021: το αποθηκευμένο workflow μπορεί να εφαρμοστεί αυτόματα στα νέα δεδομένα χωρίς δομικές αλλαγές, υπό τη μόνη προϋπόθεση ότι το νέο dataset διατηρεί συμβατή σημασιολογική αντιστοίχιση. Με τον τρόπο αυτό, ο κύκλος ζωής κλείνει και ταυτόχρονα παράγει επαναχρησιμοποιήσιμη γνώση.

Η αρχή βασίζεται σε μια καίρια διαπίστωση: εφόσον οι πολιτικές χρήσης δεδομένων (ODRL) και οι περιγραφές επεξεργασίας (DPV) είναι ήδη εκφρασμένες σε μηχανικά αναγνώσιμη μορφή, δεν υπάρχει λόγος να εξαρτάται η συμμόρφωση από ανθρώπινη ερμηνεία. Το σύστημα μπορεί και πρέπει να ερμηνεύει τους περιορισμούς και να επιβάλλει τις ενέργειες συμμόρφωσης αυτόνομα. Η μετατόπιση αυτή έχει συνέπειες πέρα από τη μηχανική: μεταφέρει την ευθύνη της νομικής συμμόρφωσης από τον χρήστη στην αρχιτεκτονική του συστήματος.

Τρεις Πυλώνες Υλοποίησης

Το Compliance-by-Design επιτυγχάνεται μέσω τριών συμπληρωματικών μηχανισμών που λειτουργούν συντονισμένα σε διαφορετικές φάσεις του κύκλου ζωής:

1. Αυτόματη Ανάλυση ODRL Policies (pre-execution). Πριν εκτελεστεί οποιοδήποτε βήμα του workflow, το σύστημα ανακτά και αναλύει σημασιολογικά τα ODRL policies κάθε εμπλεκόμενου dataset. Για κάθε `dpv:Processing` operation που πρόκειται να

- εκτελεστεί, επαληθεύεται η ύπαρξη αντίστοιχης `odrl:Permission`. Εφόσον εντοπιστεί απουσία άδειας ή ρητή `odrl:Prohibition`, η εκτέλεση αναστέλλεται και ενεργοποιείται είτε αυτόματη διαπραγμάτευση (εάν υποστηρίζεται) είτε ειδοποίηση προς τον χρήστη.
2. Αυτόματη Εισαγωγή Κόμβων Συμμόρφωσης (runtime engineering). Όταν τα ODRL policies περιλαμβάνουν `odrl:Obligation`, το σύστημα μεταφράζει την υποχρέωση σε εκτελέσιμη ενέργεια και εισάγει τους αντίστοιχους κόμβους στο workflow. Τυπικές τέτοιες περιπτώσεις είναι η εισαγωγή `dpv:delete` για υποχρεωτική διαγραφή ενδιάμεσων ευαίσθητων δεδομένων, ή `dpv:Store` για ελεγχόμενη αποθήκευση μόνο των τελικών αποτελεσμάτων. Σημαντικό είναι ότι οι κόμβοι αυτοί τοποθετούνται στη σωστή χρονική θέση της ροής π.χ. ο `dpv:delete` πάντοτε μετά τη χρήση των δεδομένων αλλά πριν την τελική αποθήκευση.
 3. Σημαιολογική Τεκμηρίωση και Audit Trail (post-execution). Κάθε εκτέλεση καταγράφεται με πλήρη μεταδεδομένα σε μορφή JSON-LD, συμπεριλαμβανομένων: των εισόδων και εξόδων κάθε βήματος, των ενεργοποιημένων πολιτικών, των αυτόματα εισαχθέντων κόμβων συμμόρφωσης και της χρονικής σήμανσης κάθε ενέργειας. Το audit log αποτελεί σημαιολογικά αναζητήσιμο τεκμήριο συμμόρφωσης, ικανό να υποστηρίξει εκ των υστέρων ελέγχους από εποπτικές αρχές (π.χ. στο πλαίσιο GDPR Άρθρου 30 - Records of Processing Activities).

Διαφοροποίηση από Συμβατικές Προσεγγίσεις

Είναι χρήσιμη η διάκριση του Compliance-by-Design από συγγενείς αλλά διακριτές έννοιες. Η αρχή Privacy-by-Design εστιάζει αποκλειστικά σε ζητήματα προστασίας ιδιωτικότητας και έχει κυρίως οργανωτικό-διαδικαστικό χαρακτήρα. Τα συμβατικά Data Governance frameworks εστιάζουν στην επιβολή πολιτικών μέσω access control mechanisms αλλά δεν παρεμβαίνουν στη λογική της αναλυτικής επεξεργασίας. Το Compliance-by-Design του προτεινόμενου πλαισίου διαφέρει σε δύο σημεία: πρώτον, επεκτείνει τη συμμόρφωση πέρα από την προστασία προσωπικών δεδομένων ώστε να καλύπτει το σύνολο των αδειοδοτικών όρων (licensing, attribution, redistribution), και δεύτερον, ενσωματώνεται απευθείας στη ροή εκτέλεσης ως ενεργός κόμβος, όχι ως εξωτερικός έλεγχος.

3.5 Από το Οικοσύστημα στην Πράξη: Εφαρμογή σε Πραγματικά Σενάρια

Η αρχιτεκτονική που περιγράφηκε στις προηγούμενες ενότητες η ενοποίηση των ανοικτών δεδομένων του TDS, η σουίτα plugins του UPCAST και το consumer Dashboard ως κεντρικός κόμβος ενορχήστρωσης αποκτά πρακτική αξία μόνο όταν δοκιμαστεί σε πραγματικές αναλυτικές προκλήσεις. Το ερώτημα που εύλογα προκύπτει είναι το εξής: μπορεί ένα τέτοιο σημαιολογικά τεκμηριωμένο και policy-aware οικοσύστημα να ανταποκριθεί στις απαιτήσεις της πραγματικής αστικής διακυβέρνησης, όπου τα δεδομένα είναι ετερογενή, οι πολιτικές χρήσης ποικίλες και οι αναλυτικές ανάγκες πολυδιάστατες;

Για να απαντηθεί το ερώτημα αυτό, το Κεφάλαιο 4 παρουσιάζει τρεις μελέτες περίπτωσης που εκτυλίσσονται εξ ολοκλήρου εντός του οικοσυστήματος UPCAST και αξιοποιούν αποκλειστικά δεδομένα του TDS. Κάθε σενάριο επιλέχθηκε προκειμένου να αναδείξει διαφορετική διάσταση του πλαισίου: το πρώτο εστιάζει σε μια «ομαλή» ροή αμιγώς ανοικτών δεδομένων, το δεύτερο εισάγει την πολυπλοκότητα της διαπραγμάτευσης πρόσβασης σε περιορισμένα δεδομένα μέσω του Negotiation Plugin και την αυτόματη επανασχεδίαση του workflow από το consumer Dashboard, και το τρίτο επιδεικνύει τη δυνατότητα διαχρονικής ανάλυσης με ενσωμάτωση χρονομεταβαλλόμενων εξωτερικών παραμέτρων. Μέσω της τριάδας αυτής επιχειρείται η πιλοτική αξιολόγηση των θεωρητικών δυνατοτήτων που

περιγράφηκαν παραπάνω, μετατρέποντας την αρχιτεκτονική του UPCAST από τεχνική περιγραφή σε λειτουργικό εργαλείο αστικής πολιτικής.

Κεφάλαιο 4ο: Μελέτες Περίπτωσης-Workflow-Driven Ανάλυση

Για την ανάδειξη του πλαισίου DPW κατασκευάστηκαν τρία αναλυτικά σενάρια, χρησιμοποιώντας datasets από το TDS. Κάθε σενάριο μοντελοποιήθηκε ως DPW μέσω του Consumer Dashboard, με αυτόματη πολιτική διαπραγμάτευση και επανασχεδίαση.

4.1 Σενάριο 1: Δείκτες Αστικού Πρασίνου

4.1.1 Πλαίσιο και Ερευνητικός Στόχος

Το πρώτο σενάριο μελέτης επικεντρώνεται στη σημασιολογικά τεκμηριωμένη ανάλυση της σχέσης μεταξύ του αστικού πρασίνου και του μόνιμου πληθυσμού του Δήμου Θεσσαλονίκης. Η αναγκαιότητα του σεναρίου προκύπτει από τη διαπίστωση ότι, παρά τη συστηματική καταγραφή τόσο των πληθυσμιακών χαρακτηριστικών όσο και της δενδροκάλυψης σε επίπεδο δήμου, οι δύο πηγές δεδομένων αξιοποιούνται συνήθως μεμονωμένα, χωρίς διασύνδεση μεταξύ τους. Ως εκ τούτου, ο κεντρικός ερευνητικός στόχος συνίσταται στην ποσοτικοποίηση της κατανομής του αστικού πρασίνου ανά κάτοικο και στην ανάδειξη της σύνθεσης του δενδροπληθυσμού ανά είδος, ώστε να υποστηριχθεί η τεκμηριωμένη χάραξη περιβαλλοντικής πολιτικής σε επίπεδο τοπικής αυτοδιοίκησης.

Η μελέτη υλοποιείται σε πλήρη ευθυγράμμιση με τη μεθοδολογία του MDAT Pilot, ακολουθώντας τον κύκλο:

Διάθεση δεδομένων → αναλυτική επεξεργασία → σημασιολογική τεκμηρίωση → αξιολόγηση

Η διαφοροποίηση από συμβατικές αναλυτικές προσεγγίσεις έγκειται στο γεγονός ότι κάθε βήμα της ροής επεξεργασίας δεν περιγράφεται μόνο διαδικαστικά (σε επίπεδο Python script), αλλά αποτυπώνεται και σημασιολογικά μέσω των DPV 2.2, ODRL 2.2 και της custom οντολογίας MDAT, καθιστώντας τη ροή policy-aware και επαναχρησιμοποιήσιμη.

4.1.2 Πηγές Δεδομένων

Αξιοποιούνται δύο ανοικτά σύνολα δεδομένων, αμφότερα διαθέσιμα μέσω του Thessaloniki Data Space υπό την άδεια Open Database License v1.0 (ODbL 1.0)

- **Urban Tree Categories (Κατηγορίες Δέντρων Δήμου Θεσσαλονίκης):** Περιλαμβάνει την ταξινόμηση των αστικών δέντρων ανά είδος, με πεδία `greek_name`, `scientific_name` και `total`. Σημασιολογικά κατηγοριοποιείται ως `dpv:EnvironmentalData` και αντιστοιχίζεται στην έννοια `mdat:TreesDistribution`.
- **Permanent Population Census 2021 (Μόνιμος Πληθυσμός Δήμου Θεσσαλονίκης):** Παρέχει τον μόνιμο πληθυσμό του Δήμου Θεσσαλονίκης (συνολικά και ανά δημοτική ενότητα, σύμφωνα με την Απογραφή 2021 της ΕΛΣΤΑΤ. Χαρακτηρίζεται ως `dpv:DemographicData` και αντιστοιχίζεται στην έννοια `mdat:PopulationDistribution`.

Και τα δύο σύνολα διατίθενται σε μορφή Excel (.xlsx), γεγονός που καθιστά αναγκαίο ένα διακριτό στάδιο καθαρισμού και κανονικοποίησης πριν από οποιαδήποτε αναλυτική επεξεργασία.

4.1.3 Ρόλοι Οντολογίας

Σύμφωνα με το DPV 2.2, στο σενάριο αναγνωρίζονται τρεις διακριτοί ρόλοι, καθένας με σαφώς οριοθετημένες αρμοδιότητες:

- Data Provider (dpv:DataSource, dpv:DataController): Παρέχει τα επίσημα σύνολα δεδομένων (πληθυσμιακή απογραφή και κατηγορίες δέντρων) και είναι υπεύθυνος για την αρχική τους διάθεση υπό τους όρους της αντίστοιχης άδειας.
- Data Analyst (dpv:DataProcessor): Εκτελεί την τεχνική επεξεργασία των δεδομένων, υλοποιεί τα Python scripts, παράγει τους δείκτες και τις οπτικοποιήσεις.
- Researcher (dpv:Recipient, dpv:DataUser): Ερμηνεύει τα αποτελέσματα, αξιολογεί την περιβαλλοντική επίπτωση και τα μεταφράζει σε εισηγήσεις πολιτικής.

Η διάκριση αυτή είναι ουσιώδης για την εφαρμογή της αρχής Compliance-by-Design, καθώς κάθε ρόλος συσχετίζεται με συγκεκριμένες ODRL πολιτικές χρήσης που καθορίζουν τα επιτρεπτά actions επί των δεδομένων.

4.1.4 Σημασιολογική Αντιστοίχιση Διεργασιών

Η ροή επεξεργασίας απαρτίζεται από τέσσερα διακριτά βήματα, καθένα εκ των οποίων μοντελοποιείται ως dpv:Process και αντιστοιχίζεται τόσο σε κλάσεις DPV και actions ODRL όσο και σε εξειδικευμένες έννοιες του namespace mdat:. Ο Πίνακας 4.1 παρουσιάζει την πλήρη αντιστοίχιση:

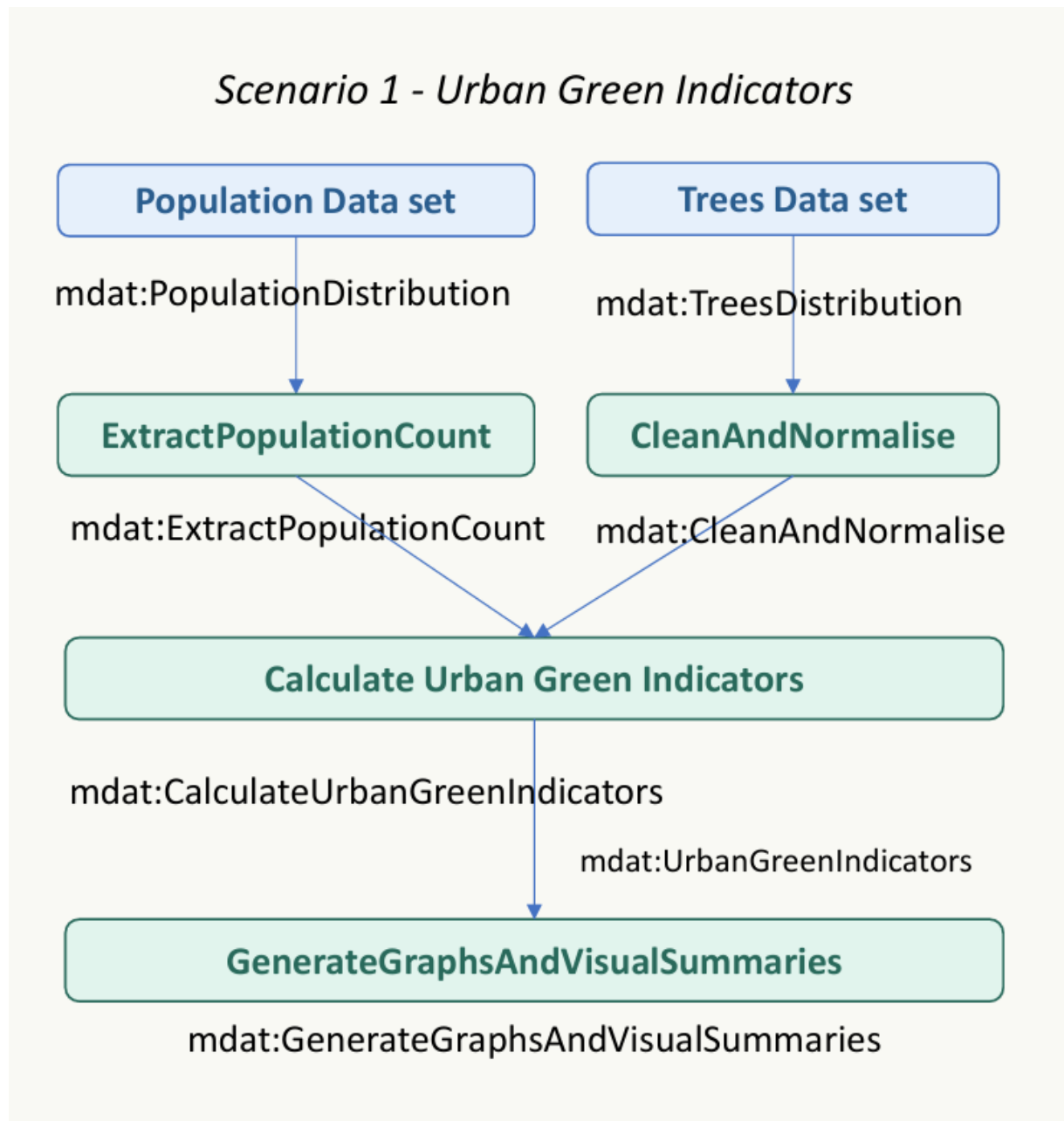
Βήμα	Κόμβος	DPV	ODRL	MDAT Concept	Ρόλος
1	ExtractPopulationCount	dpv:Collect, dpv:Derive	odrl:use	mdat:ExtractPopulationCount → mdat:PopulationCount	Data Analyst
2	CleanAndNormalise	dpv:Transform	odrl:modify	mdat:CleanAndNormalise	Data Analyst
3	CalculateUrbanGreenIndicators	dpv:Derive, dpv:Aggregate	odrl:derive	mdat:CalculateUrbanGreenIndicators → mdat:UrbanGreenIndicators	Data Analyst
4	GenerateGraphsAndVisualSummaries	dpv:Visualise, dpv:Analyse	odrl:display, reproduce	mdat:GenerateGraphsAndVisualSummaries	Researcher

Πίνακας 4.1: Σενάριο 1 - Αποσύνθεση διεργασιών και σημασιολογική αντιστοίχιση

Αξίζει να σημειωθεί ότι οι έννοιες του namespace mdat: εισήχθησαν στις περιπτώσεις όπου οι γενικοί όροι των DPV/ODRL δεν επαρκούσαν για την περιγραφή της εξειδικευμένης σημασιολογίας του πεδίου εφαρμογής. Για παράδειγμα, ενώ το dpv:Derive περιγράφει γενικά τη διαδικασία παραγωγής νέων δεδομένων, η έννοια mdat:CalculateUrbanGreenIndicators αποδίδει με ακρίβεια το συγκεκριμένο υπολογισμό δεικτών αστικού πρασίνου που εκτελείται στο τρίτο βήμα.

4.1.5 Ροή Επεξεργασίας Δεδομένων

Η αρχιτεκτονική του workflow, όπως αποτυπώνεται στο Σχήμα 4.1, ακολουθεί ένα υβριδικό μοτίβο παράλληλης και σειριακής εκτέλεσης.



Σχήμα 4.1: Data Processing Workflow - Σενάριο 1: Δείκτες Αστικού Πρασίνου

Συγκεκριμένα:

1. **ExtractPopulationCount**: Δέχεται ως είσοδο το dataset `mdat:PopulationDistribution` και εξάγει το πληθυσμιακό μέγεθος ανά δημοτική ενότητα, παράγοντας την οντότητα `mdat:PopulationCount`. Σημασιολογικά η ενέργεια αντιστοιχίζεται σε `dpv:Collect` (συλλογή του υποσυνόλου) σε συνδυασμό με `dpv:Derive` (παραγωγή του τελικού αριθμητικού δείκτη).
2. **CleanAndNormalise**: Λειτουργεί παράλληλα με το πρώτο βήμα στη ροή του dataset δέντρων (`mdat:TreesDistribution`). Περιλαμβάνει την εξυγίανση των επιστημονικών

και ελληνικών ονομασιών των ειδών, την ομογενοποίηση των μονάδων και την απομάκρυνση ελλιπών εγγραφών. Η αντιστοίχιση με `dpv:Transform` και `odrl:modify` αντικατοπτρίζει τον μετασχηματιστικό χαρακτήρα της λειτουργίας χωρίς παραγωγή ουσιαστικά νέων δεδομένων.

3. `CalculateUrbanGreenIndicators`: Αποτελεί το αναλυτικό επίκεντρο του σεναρίου. Συγκλίνουν σε αυτό οι έξοδοι των δύο προηγούμενων βημάτων και υπολογίζονται οι δείκτες *δέντρα ανά κάτοικο*, τόσο συνολικά όσο και ανά είδος, παράγοντας την οντότητα `mdat:UrbanGreenIndicators`. Ο διπλός χαρακτηρισμός `dpv:Derive` + `dpv:Aggregate` αποτυπώνει τόσο τη συναθροιστική όσο και την παραγωγική διάσταση της ενέργειας.
4. `GenerateGraphsAndVisualSummaries`: Παράγει τα τελικά παραδοτέα οπτικοποίησης γραφήματα δέντρων ανά κάτοικο, συγκριτικά διαγράμματα συνολικού πληθυσμού έναντι ανά είδος και ραβδογράμματα απόλυτων αριθμών, τα οποία απευθύνονται στους υπεύθυνους χάραξης αστικής πολιτικής. Η αντιστοίχιση με `odrl:display` και `odrl:reproduce` υποδηλώνει τα επιτρεπόμενα δικαιώματα προβολής και αναπαραγωγής των οπτικοποιήσεων.

Τα βήματα 1 (`ExtractPopulationCount`) και 2 (`CleanAndNormalise`) εκτελούνται παράλληλα στις αντίστοιχες ροές εισόδου. Οι έξοδοί τους συγκλίνουν στο `CalculateUrbanGreenIndicators`, που υπολογίζει τον λόγο δέντρων ανά κάτοικο (`mdat:UrbanGreenIndicators`). Το τελευταίο βήμα παράγει οπτικές περιλήψεις για τους υπεύθυνους αστικής πολιτικής.

Η παράλληλη εκτέλεση των βημάτων 1 και 2, ακολουθούμενη από το σημείο σύγκλισης στο βήμα 3, επιτρέπει την αποδοτική διαχείριση των δύο ετερογενών ροών εισόδου και την ανεξάρτητη επικαιροποίηση του καθενός από τα δύο datasets, χωρίς να απαιτείται επανεκτέλεση ολόκληρου του workflow.

4.1.6 Δομή Παραδοτέων και Αποτελέσματα Εκτέλεσης

4.1.6.1 Δομή Παραδοτέων κατά την εκτέλεση του workflow

Τα αναμενόμενα παραδοτέα που θα λαμβάνονται σε κάθε εκτέλεση του workflow θα περιλαμβάνουν τους παρακάτω τύπους δεδομένων:

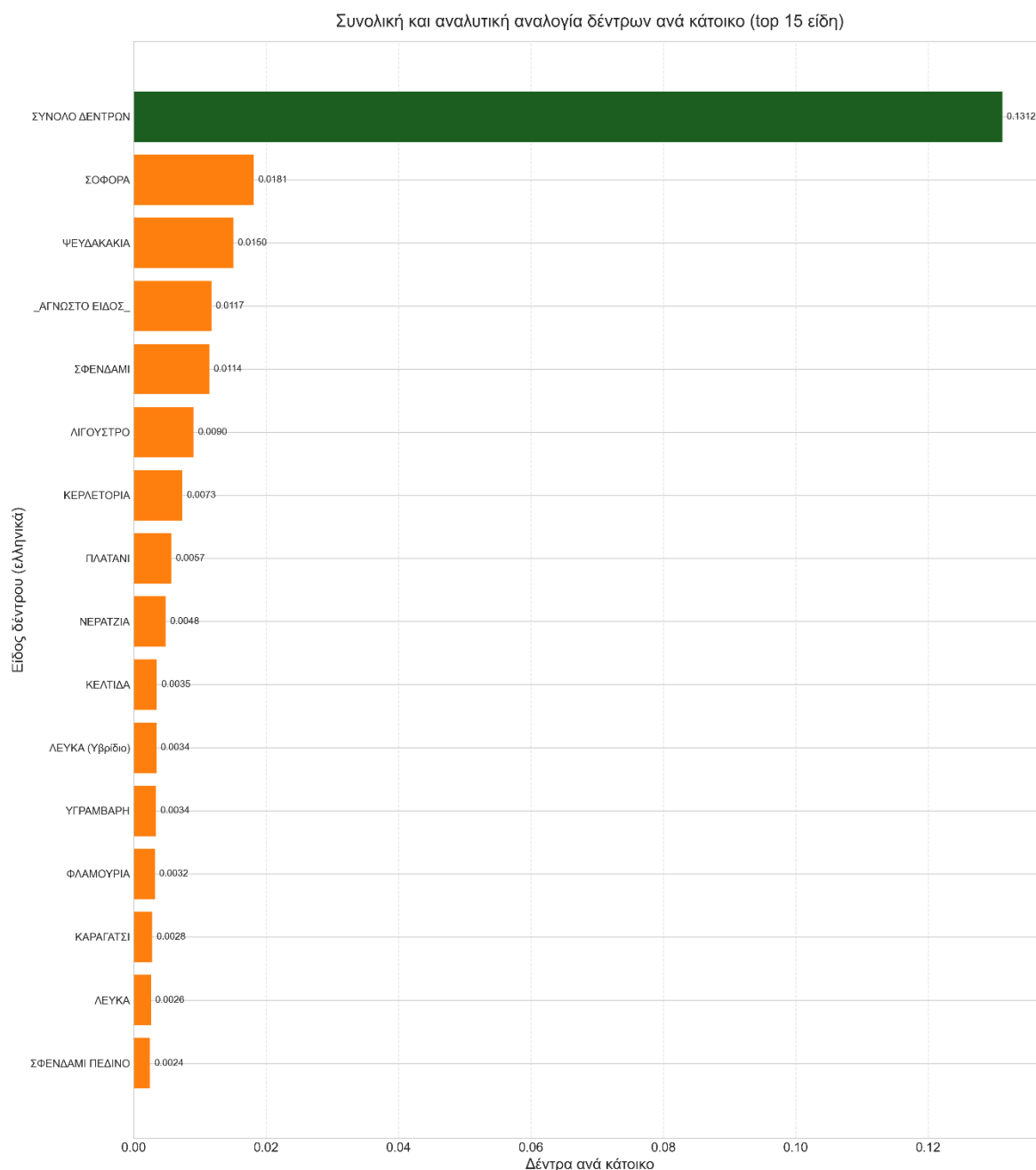
- Το αρχείο `trees_per_citizen.xlsx` με τους υπολογισθέντες δείκτες ανά είδος και συνολικά (`dpv:DerivedData`, `dpv:StatisticalData`).
- Σύνολο γραφημάτων σε μορφή PNG (`dpv:VisualisationData`), συμπεριλαμβανομένων διαγραμμάτων δέντρων ανά κάτοικο, συγκριτικών αναπαραστάσεων ανά είδος και ραβδογραμμάτων απόλυτων τιμών.

Η πλήρης υλοποίηση είναι διαθέσιμη στο δημόσιο αποθετήριο MDAT-scenario-thessaloniki-green, με τη δομή που υπαγορεύει η μεθοδολογία του MDAT Pilot: 1_Scenario_Description, 2_Data_Specification, 3_Code, 4_Outputs και 5_Metadata_Policies.

4.1.6.2 Αποτελέσματα και Οπτικοποιήσεις Σεναρίου 1

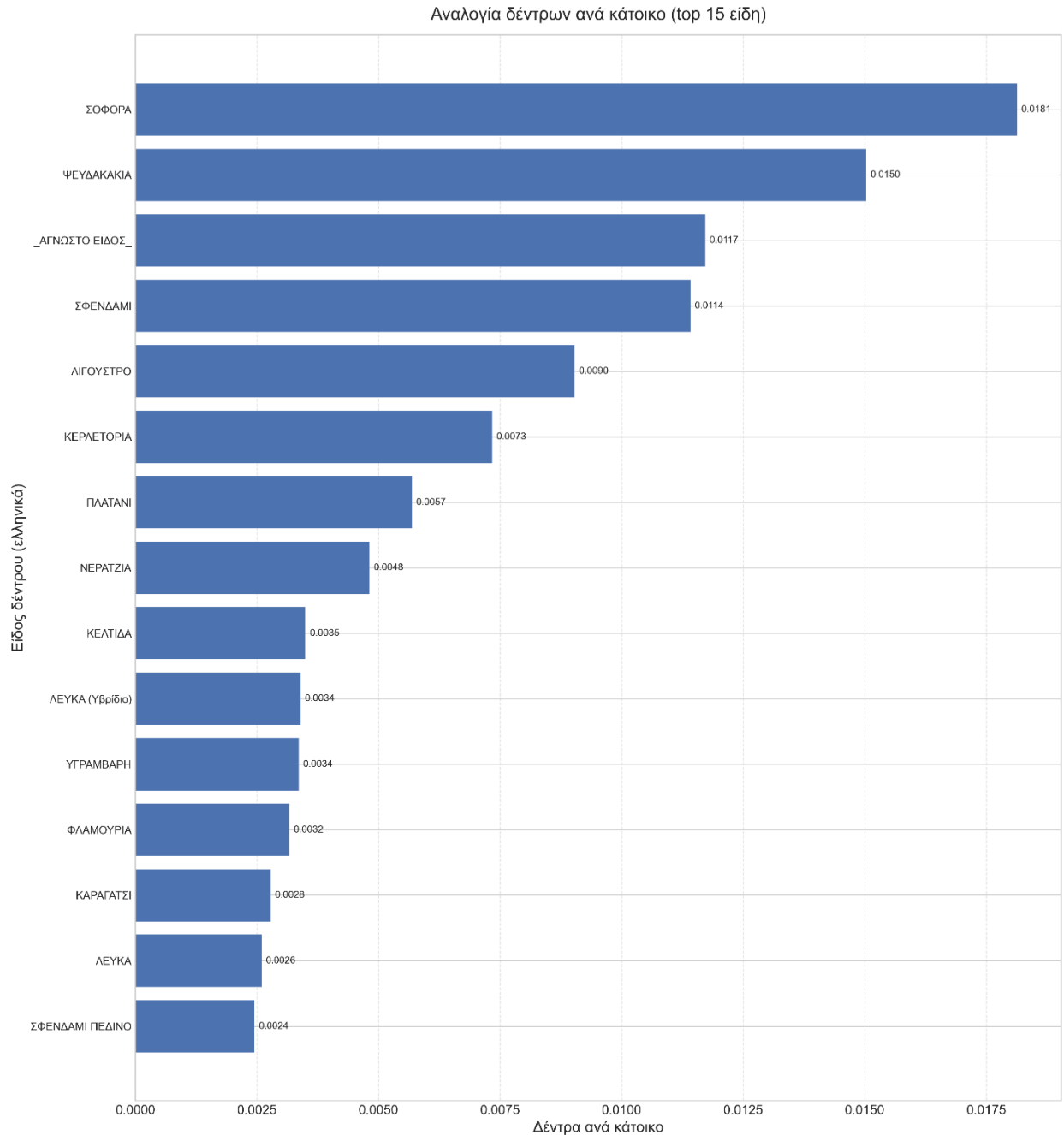
Η εκτέλεση της διεργασίας `mdat:ExtractPopulationCount` επί της Απογραφής 2021 απέδωσε μόνιμο πληθυσμό 319.045 κατοίκων για τον Δήμο Θεσσαλονίκης, ενώ η `mdat:CleanAndNormalise` παρήγαγε εξυγιανθέν σύνολο 90 ειδών δέντρων με πλήρη ονοματολογία (ελληνική και επιστημονική). Επί αυτών των δύο εξόδων εφαρμόστηκε η `mdat:CalculateUrbanGreenIndicators`, η οποία απέδωσε συνολικό λόγο 0,1312 δέντρα ανά κάτοικο - δηλαδή περίπου ένα δέντρο ανά 7,6 κατοίκους.

Το Σχήμα 4.2 συνοψίζει συγκριτικά τη συνολική αναλογία έναντι των επιμέρους ειδών, αναδεικνύοντας τη δραστική διαφορά μεγέθους μεταξύ του αθροιστικού δείκτη και των μεμονωμένων συνεισφορών.



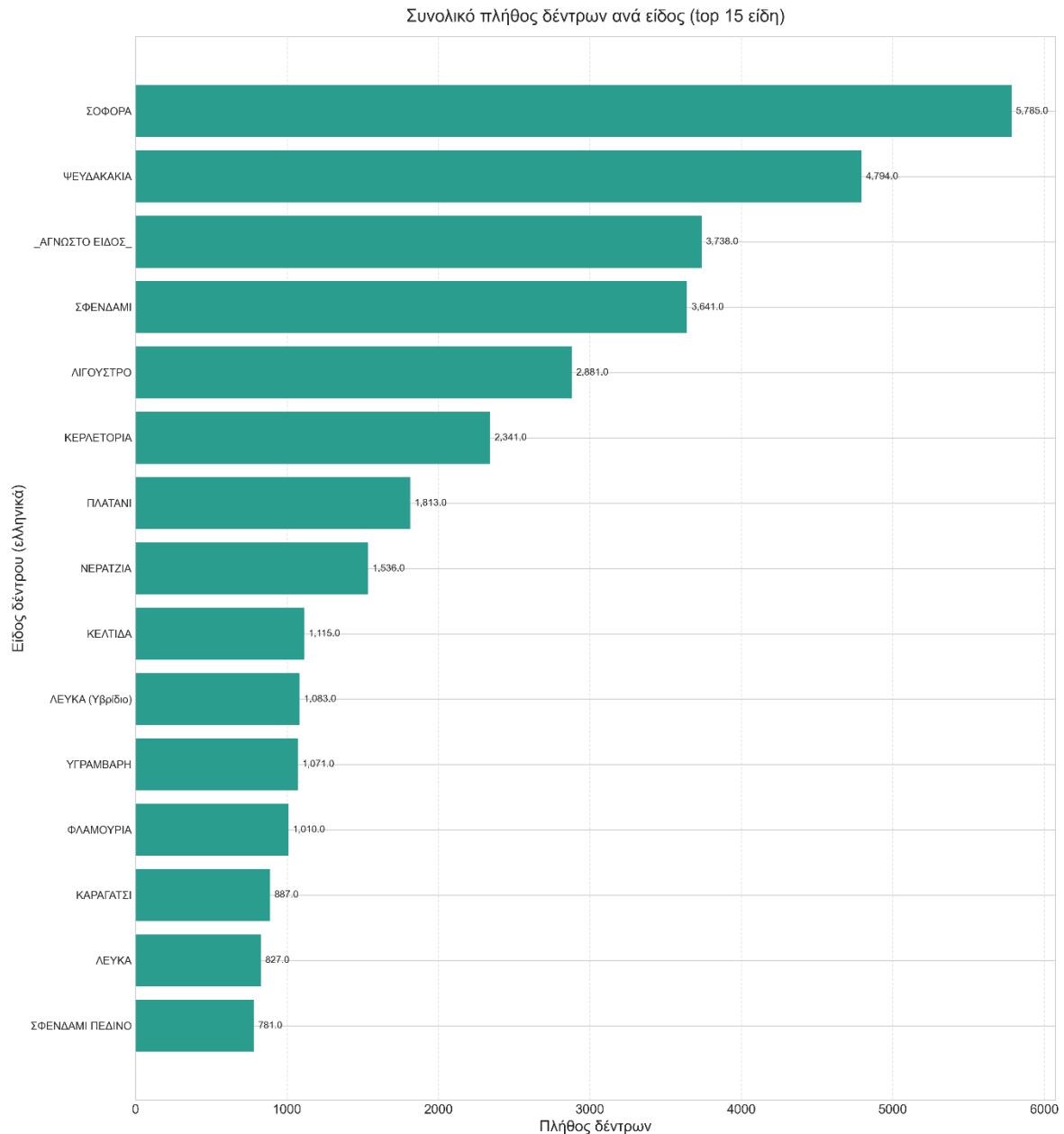
Σχήμα 4.2: Συνολική και αναλυτική αναλογία δέντρων ανά κάτοικο για τα 15 πολυπληθέστερα είδη - Δήμος Θεσσαλονίκης (πληθυσμός αναφοράς: 319.045 κάτοικοι).

Σε ένα δεύτερο επίπεδο ανάλυσης, το **Σχήμα 4.3** εστιάζει αποκλειστικά στην ανά είδος αναλογία, αποκαλύπτοντας την επικράτηση της *Σοφώρας* (0,0181) και της *Ψευδακακίας* (0,0150). Αξιοσημείωτη είναι η παρουσία της κατηγορίας «Άγνωστο Είδος» (0,0117) στην τρίτη θέση, η οποία υποδεικνύει έλλειμμα στην αρχική καταχώριση των δεδομένων και αποτελεί εύρημα με αξία για τον πάροχο των δεδομένων.



Σχήμα 4.3: Κατά είδος αναλογία δέντρων ανά κάτοικο (top 15) - αποτέλεσμα της διεργασίας *mdat:CalculateUrbanGreenIndicators*.

Τέλος, το Σχήμα 4.4 παρουσιάζει τις απόλυτες τιμές πληθυσμού δέντρων ανά είδος. Τα 15 πολυπληθέστερα είδη αθροίζουν 29.302 δέντρα επί συνολικού δενδροπληθυσμού 41.866, ποσοστό που αγγίζει το 70% και αναδεικνύει τη σαφή συγκέντρωση της δενδροκάλυψης σε περιορισμένο αριθμό ειδών, εύρημα με ενδιαφέρον για ζητήματα βιοποικιλότητας και ανθεκτικότητας του αστικού οικοσυστήματος.



Σχήμα 4.4: Απόλυτο πλήθος δέντρων ανά είδος (top 15) - Δήμος Θεσσαλονίκης, dataset «Κατηγορίες Δέντρων» (ODbL 1.0).

Συνολικά, τα παραπάνω αποτελέσματα δεν αποτελούν αυτοσκοπό· η αξία τους έγκειται στο ότι παράγονται από μια πλήρως τεκμηριωμένη και policy-aware ροή, στην οποία κάθε ενδιάμεσο και τελικό artifact είναι σημασιολογικά συνδεδεμένο με τις αντίστοιχες DPV/ODRL/MDAT έννοιες.

4.1.7 Συμπεράσματα του Σεναρίου

Το πρώτο σενάριο παρουσιάζει εμπράκτως ότι η ενσωμάτωση των DPV/ODRL με μια εξειδικευμένη οντολογία πεδίου (MDAT) επιτρέπει τη μετατροπή μιας συμβατικής αναλυτικής διεργασίας σε ένα πλήρως τεκμηριωμένο, policy-aware και επαναχρησιμοποιήσιμο workflow. Η σημασιολογική διαλειτουργικότητα που επιτυγχάνεται δεν αποτελεί απλώς θεωρητικό πλεονέκτημα, αλλά μεταφράζεται σε πρακτικές δυνατότητες ιχνηλασιμότητας, ελέγχου

συμμόρφωσης και αυτοματοποιημένης επιβολής περιορισμών χρήσης, στοιχεία απαραίτητα για την ένταξη της ροής σε ευρύτερα Data Space περιβάλλοντα.

Πέραν της μεθοδολογικής συνεισφοράς, το σενάριο ανέδειξε και ευρήματα ουσιαστικής αξίας: ο μέσος δείκτης 0,1312 δέντρων ανά κάτοικο τοποθετεί τον Δήμο Θεσσαλονίκης σε χαμηλά επίπεδα συγκριτικά με ευρωπαϊκούς δείκτες αναφοράς, ενώ η συγκέντρωση του 70% του δενδροπληθυσμού σε 15 μόνο είδη εγείρει ζητήματα βιοποικιλότητας και ανθεκτικότητας. Επιπλέον, η ύπαρξη της κατηγορίας «Άγνωστο Είδος» (3.738 εγγραφές, ≈9% του συνόλου) αποτελεί ποιοτικό εύρημα ανά-ροής (data quality feedback) προς τον πάροχο των δεδομένων, υποδεικνύοντας την ανάγκη συστηματικής ταυτοποίησης.

4.2 Σενάριο 2: Δείκτες Ποιότητας Αέρα και Πληθυσμιακής Έκθεσης

4.2.1 Πλαίσιο και Ερευνητικός Στόχος

Το δεύτερο σενάριο μελέτης διερευνά τη σχέση μεταξύ της ατμοσφαιρικής ρύπανσης και του μόνιμου πληθυσμού στις δημοτικές κοινότητες του Δήμου Θεσσαλονίκης, με κεντρικό στόχο την ποσοτικοποίηση της περιβαλλοντικής έκθεσης ανά κάτοικο και τον εντοπισμό περιοχών με αυξημένη επιβάρυνση. Σε αντιδιαστολή με το πρώτο σενάριο, όπου αξιοποιήθηκαν αποκλειστικά ανοικτά δεδομένα ομοιογενούς καθεστώτος, το παρόν σενάριο εισάγει μια ποιοτικά διαφορετική διάσταση: τη διαπραγμάτευση πρόσβασης σε δεδομένα περιορισμένης διάθεσης μέσω του UPCAST Negotiation Plugin και την αυτόματη επανασχεδίαση της ροής επεξεργασίας από το Consumer Dashboard σε απόκριση των αντίστοιχων ODRL πολιτικών χρήσης.

Η αναγκαιότητα του σεναρίου προκύπτει από τη διαπίστωση ότι τα δημοσίως διαθέσιμα πληθυσμιακά δεδομένα συχνά δεν διαθέτουν την απαιτούμενη χωρική ανάλυση για την ακριβή συσχέτιση δεικτών ρύπανσης με τον εκτεθειμένο πληθυσμό σε επίπεδο δημοτικής κοινότητας ή σταθμού μέτρησης. Η αξιοποίηση δεδομένων υψηλότερης ανάλυσης απαιτεί ελεγχόμενη πρόσβαση υπό συγκεκριμένους όρους χρήσης, γεγονός που καθιστά το σενάριο ιδανικό πεδίο επίδειξης των μηχανισμών Compliance-by-Design του MDAT Pilot.

Η ροή επεξεργασίας ακολουθεί τον επεκταμένο κύκλο:

Διαπραγμάτευση πρόσβασης → διάθεση δεδομένων → αυτόματη επανασχεδίαση workflow → αναλυτική επεξεργασία → σημασιολογική τεκμηρίωση → αξιολόγηση συμμόρφωσης

4.2.2 Πηγές Δεδομένων

Στο σενάριο αξιοποιούνται τρία σύνολα δεδομένων εισόδου με διαφορετικό καθεστώς πρόσβασης, η σύζευξη των οποίων αναδεικνύει τη δυνατότητα του πλαισίου να ενοποιεί ετερογενείς πηγές υπό διαφορετικά αδειοδοτικά καθεστώτα:

- D1 - Air Quality Measurements (2010–2013): Μετρήσεις ατμοσφαιρικής ρύπανσης από το Δημοτικό Δίκτυο Παρακολούθησης της Θεσσαλονίκης, με τιμές για επτά ρύπους (CO, NO, NO₂, O₃, PM₁₀, PM_{2.5}, SO₂). Σημασιολογικά κατηγοριοποιείται ως mdat:EnvironmentalData / dpv:NonPersonalData υπό άδεια ODbL 1.0 και διατίθεται δημοσίως.
- D2 - Permanent Population Census 2011 (public): Πληθυσμός ανά δημοτική κοινότητα σύμφωνα με την Απογραφή 2011 της ΕΛΣΤΑΤ. Χαρακτηρίζεται ως mdat:DemographicData υπό άδεια ODbL 1.0.

- D3 - Extended Population Dataset (negotiated access): Επεκτεταμένο σύνολο δημογραφικών δεδομένων με υπο-δημοτική κατάτμηση για τη μητροπολιτική περιοχή Θεσσαλονίκης. Χαρακτηρίζεται ως `mdat:ExtendedPopulationData` και ταυτόχρονα ως `dpv:RestrictedData`, διατιθέμενο υπό καθεστώς ελεγχόμενης πρόσβασης (*Controlled Access*) κατόπιν επιτυχούς διαπραγμάτευσης. Σημειώνεται ότι το D3 αξιοποιείται στο παρόν σενάριο πρωτίστως ως μέσο επίδειξης (demonstration dataset) του Negotiation Plugin και των μηχανισμών Compliance-by-Design του Consumer Dashboard, εντός του πιλοτικού πλαισίου του MDAT.

Επιπλέον, ως παραγόμενο σύνολο (D4) αναγνωρίζεται το `atmospheric_analysis_thessaloniki.xlsx` μαζί με τα συνοδευτικά γραφήματα PNG, χαρακτηρισμένο ως `dpv:DerivedData` / `mdat:ExposureIndicator` υπό άδεια CC BY-NC 4.0.

Η συνύπαρξη ανοικτών (D1, D2) και περιορισμένα διαθέσιμων (D3) δεδομένων στο ίδιο workflow αποτελεί τη βασική πρόκληση και ταυτόχρονα τη βασική καινοτομία του σεναρίου.

4.2.3 Ρόλοι Οντολογίας

Οι ρόλοι του σεναρίου επεκτείνουν τον ορισμό του DPV πέρα από την απλή επεξεργασία, ώστε να συμπεριλάβουν τη διαχείριση δικαιωμάτων πρόσβασης και τη διαπραγμάτευση όρων χρήσης:

- Data Provider (`dpv:DataController`, `dpv:Authority`, `dpv:AccessControlMethod`): Διαχειρίζεται τα επίσημα σύνολα δεδομένων, καθορίζει τις πολιτικές πρόσβασης και διεκπεραιώνει τις διαπραγματεύσεις χρήσης. Σε αντίθεση με το παθητικό προφίλ του Σεναρίου 1, εδώ ο ρόλος του είναι ενεργός κατά τη φάση διαπραγμάτευσης.
- Researcher (`dpv:Recipient` με σκοπό `dpv:ResearchAndDevelopment`, `odrl:assignee`): Εκκινεί τη διαπραγμάτευση, τεκμηριώνει την επιστημονική αναγκαιότητα πρόσβασης στο `mdat:ExtendedPopulationData` και ερμηνεύει τα τελικά αποτελέσματα προς υποστήριξη περιβαλλοντικής πολιτικής.
- Data Analyst (`dpv:DataProcessor`, `odrl:operator`): Εκτελεί τους μετασχηματισμούς, τους υπολογισμούς και τις οπτικοποιήσεις, λειτουργώντας εντός των πολιτικών χρήσης που έχουν συμφωνηθεί.

4.2.4 Διαπραγμάτευση και Σταδιακή Πρόσβαση

Κεντρικό στοιχείο διαφοροποίησης του σεναρίου αποτελεί η διαδικασία διαπραγμάτευσης πρόσβασης, η οποία διεξάγεται μέσω του UPCAST Negotiation Plugin. Κατά την αναλυτική φάση, ο Researcher διαπίστωσε ότι το δημόσιο dataset D2 δεν διέθετε την απαιτούμενη χωρική ανάλυση για τη σωστή αντιστοίχιση πληθυσμιακών μεγεθών με τους επιμέρους σταθμούς μέτρησης. Προς επίλυση του ζητήματος, εκκίνησε επίσημη διαδικασία διαπραγμάτευσης, η οποία περιλάμβανε τρία διακριτά βήματα:

1. Τεκμηρίωση επιστημονικής αναγκαιότητας από τον Researcher προς τον Data Provider, με ρητή αναφορά στον σκοπό χρήσης (`dpv:ResearchAndDevelopment`) και στις απαιτήσεις χωρικής ανάλυσης.
2. Αξιολόγηση του αιτήματος από τον Data Provider και διαμόρφωση πολιτικής σταδιακής πρόσβασης (*graded-access policy*) με όρους ελεγχόμενης επαναχρησιμοποίησης, προστασίας ιδιωτικότητας και υποχρεωτικής παραπομπής.
3. Σύναψη συμφωνίας, η οποία μοντελοποιείται σημασιολογικά ως `dpv:NegotiateContract` σε συνδυασμό με την εξειδικευμένη έννοια

mdat:NegotiatedAccessPolicy. Με βάση τη συμφωνία αυτή, παραχωρήθηκε πρόσβαση στο dataset D3.

Η πολιτική που συνοδεύει το D3 περιλαμβάνει αυστηρούς όρους, μεταξύ των οποίων ξεχωρίζουν η υποχρεωτική διαγραφή των ενδιάμεσων ευαίσθητων δεδομένων μετά την ολοκλήρωση της επεξεργασίας και η διατήρηση αποκλειστικά του τελικού συγκεντρωτικού αποτελέσματος.

4.2.5 Αυτόματη Επανασχεδίαση Workflow

Σε απόκριση της ODRL πολιτικής του dataset D3, το Consumer Dashboard ανέλυσε αυτόματα τους περιορισμούς χρήσης και προέβη σε δυναμική επανασχεδίαση του workflow, εισάγοντας δύο επιπλέον κόμβους-συστήματος που δεν προβλέπονταν στην αρχική σχεδίαση από τον Data Analyst:

- Έναν κόμβο dpv:Erase με ODRL action odrl:delete, τοποθετημένο αμέσως μετά την παραγωγή των συγκεντρωτικών δεικτών, ο οποίος εξασφαλίζει την υποχρεωτική διαγραφή των ενδιάμεσων ευαίσθητων δεδομένων.
- Έναν κόμβο dpv:Store με ODRL action odrl:reproduce, στο τέλος της ροής, ο οποίος επιτρέπει τη διατήρηση αποκλειστικά της τελικής αναφοράς αποτελεσμάτων.

Η δυνατότητα αυτή της αυτόματης ένθεσης κόμβων αποτελεί εφαρμογή της αρχής Compliance-by-Design: ο Researcher και ο Data Analyst δεν χρειάζεται να γνωρίζουν εκ των προτέρων τις λεπτομέρειες της πολιτικής χρήσης, καθώς το σύστημα ερμηνεύει σημασιολογικά τους περιορισμούς και επιβάλλει αυτόματα τις απαιτούμενες ενέργειες συμμόρφωσης. Με τον τρόπο αυτό, η τήρηση των όρων αδειοδότησης μπορεί να ενσωματωθεί ως εκτελέσιμη διαδικασία μέσω του συστήματος.

4.2.6 Σημασιολογική Αντιστοίχιση Διεργασιών

Η ροή επεξεργασίας απαρτίζεται από επτά διακριτά βήματα, εκ των οποίων δύο (4* και 7*) εισάγονται αυτόματα από το Consumer Dashboard. Ο Πίνακας 5.2 παρουσιάζει την πλήρη αντιστοίχιση :

Βήμα	Κόμβος	DPV	ODRL	MDAT Concept	Ρόλος
1	AggregateAndComputeMeanPollutantLevels	dpv:Aggregate, dpv:Derive	odrl:derive	mdat:AggregateAndComputeMeanPollutantLevels	Data Analyst
2	CleanAndNormalise	dpv:Transform	odrl:modify	mdat:CleanAndNormalise	Data Analyst
3	ComputePollutantPerCapitaRatios	dpv:Combine, dpv:Derive	odrl:aggregate	mdat:ComputePollutantPerCapitaRatios	Data Analyst
4*	dpv:delete (αυτόματο)	dpv:Erase	odrl:delete	-	Σύστημα

5	AssessComplianceWithEULimits	dpv:Assess	odrl:analyze	mdat:AssessComplianceWithEULimits	Data Analyst
6	GenerateGraphsAndVisualSummaries	dpv:Visualise, dpv:Analyse	odrl:display, reproduce	mdat:GenerateGraphsAndVisualSummaries	Researcher
7*	dpv:store (αυτόματο)	dpv:Store	odrl:reproduce	—	Σύστημα

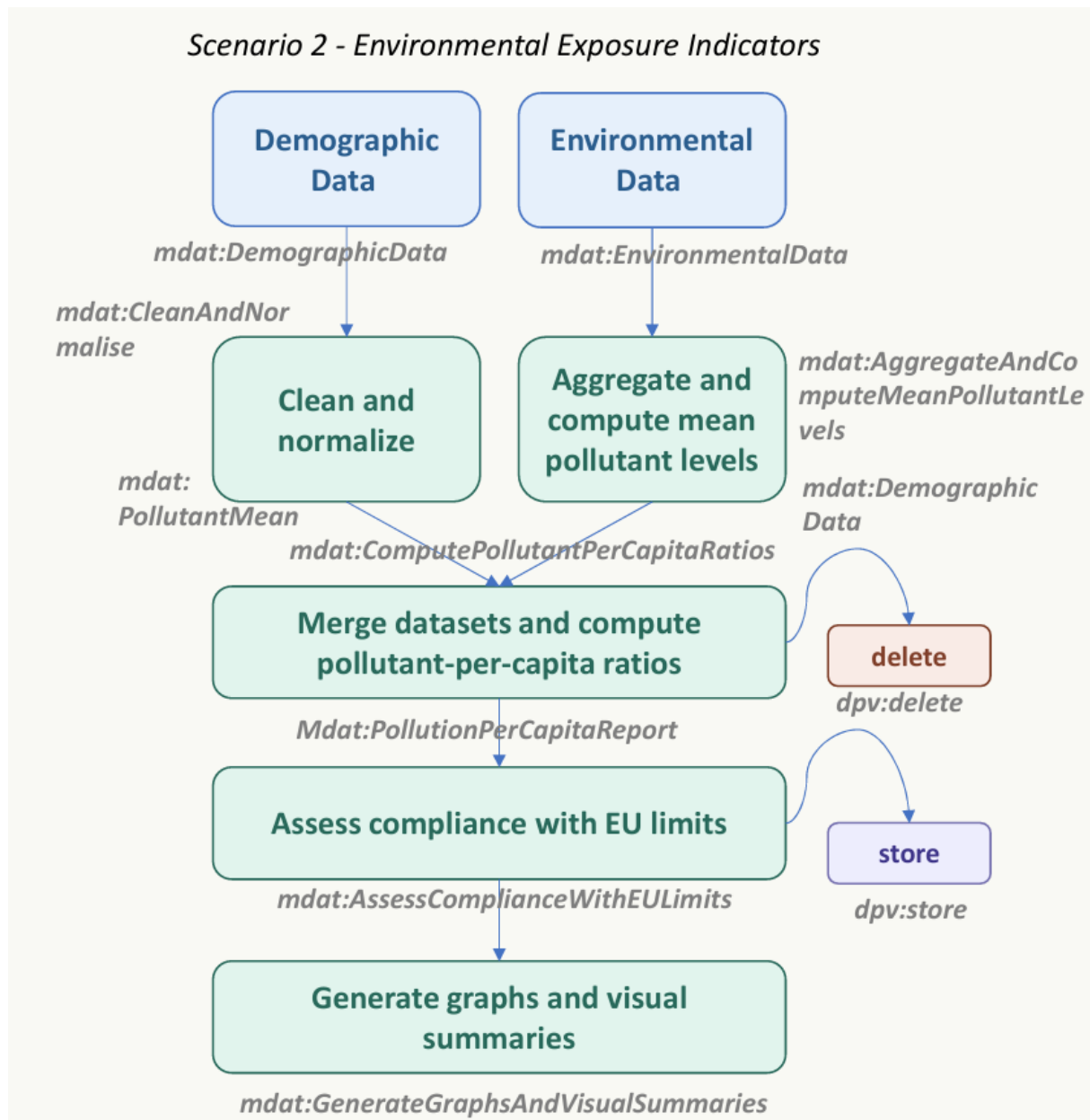
Πίνακας 4.2: Σενάριο 2 - Αποσύνθεση διεργασιών και σημασιολογική αντιστοίχιση. Με αστερίσκο (*) σημειώνονται τα βήματα που εισάγονται αυτόματα από το ABOVO Dashboard σε απόκριση των ODRL πολιτικών του dataset περιορισμένης πρόσβασης.*

Όπως και στο πρώτο σενάριο, οι έννοιες του namespace mdat: εισήχθησαν στις περιπτώσεις όπου οι γενικοί όροι των DPV/ODRL δεν επαρκούσαν για την περιγραφή της εξειδικευμένης σημασιολογίας του πεδίου. Ιδιαίτερη αξία έχει η έννοια mdat:NegotiatedAccessPolicy, η οποία αποτυπώνει με ακρίβεια τη φύση της συμφωνίας σταδιακής πρόσβασης που υπερβαίνει την απλή αδειοδότηση και ενσωματώνει τη διαπραγματευτική διάσταση.

4.2.7 Ροή Επεξεργασίας Δεδομένων

Η αρχιτεκτονική του workflow, όπως αποτυπώνεται στο Σχήμα 4.5, οργανώνεται σε τρεις λειτουργικές φάσεις:

- **Φάση Α - Διαπραγμάτευση και Συλλογή Δεδομένων.** Η ροή ξεκινά με τη διαπραγμάτευση πρόσβασης μεταξύ Researcher και Data Provider μέσω του UPCAST Negotiation Plugin, η οποία οδηγεί στην παραχώρηση του D3. Παράλληλα συλλέγονται τα ανοικτά δεδομένα D1 και D2.
- **Φάση Β - Επεξεργασία και Ανάλυση.** Τα δύο ρεύματα εισόδου διέρχονται από την εξυγίανση και κανονικοποίηση (βήμα 2), που εξασφαλίζει την ομογενοποίηση των μονάδων μέτρησης (mg/m^3 έναντι $\mu\text{g}/\text{m}^3$) και την απομάκρυνση ασυνεπών εγγραφών. Ακολουθεί η υπολογιστική σύνθεση των μέσων συγκεντρώσεων ρύπων ανά σταθμό για την περίοδο 2010-2013 (βήμα 1), και έπεται ο συνδυασμός με τα δημογραφικά δεδομένα για τον υπολογισμό των δεικτών έκθεσης ανά κάτοικο (βήμα 3). Σε αυτή τη φάση, αμέσως μετά την παραγωγή των συγκεντρωτικών δεικτών, ενεργοποιείται αυτόματα ο κόμβος dpv:delete (βήμα 4*), εξασφαλίζοντας τη διαγραφή των ενδιάμεσων ευαίσθητων δεδομένων του D3.
- **Φάση Γ - Αξιολόγηση και Παρουσίαση.** Το βήμα 5 αξιολογεί τη συμμόρφωση των μετρήσεων με τα όρια της Ευρωπαϊκής Ένωσης, παράγοντας έναν δείκτη dpv:ComplianceStatus ανά ρύπο και ανά δημοτική κοινότητα. Το βήμα 6 παράγει τις τελικές οπτικοποιήσεις και ο κόμβος dpv:store (βήμα 7*) διατηρεί αποκλειστικά τα τελικά αναφερόμενα παραδοτέα.



Σχήμα 4.5: Data Processing Workflow - Σενάριο 2: Δείκτες Ποιότητας Αέρα και Πληθυσμιακής Έκθεσης, με τους αυτόματους κόμβους *dpv:delete* και *dpv:store* που εισήχθησαν από το ABOVO Dashboard σε απόκριση των ODRL πολιτικών του dataset περιορισμένης πρόσβασης.

4.2.8 Δομή Παραδοτέων και Αποτελέσματα Εκτέλεσης

4.2.8.1 Δομή Παραδοτέων κατά την εκτέλεση του workflow

Τα αναμενόμενα παραδοτέα που θα λαμβάνονται σε κάθε εκτέλεση του workflow θα περιλαμβάνουν τους παρακάτω τύπους δεδομένων:

- Αρχείο `atmospheric_analysis_thessaloniki.xlsx` με τους υπολογισμένους δείκτες έκθεσης και τα αποτελέσματα ελέγχου συμμόρφωσης (*dpv:DerivedData*, *mdat:ExposureIndicator*).

- Σύνολο επτά γραφημάτων σε μορφή PNG (dpv:VisualisationData), ένα ανά ρύπο: CO_by_district.png, NO_by_district.png, NO2_by_district.png, O3_by_district.png, PM10_by_district.png, PM2.5_by_district.png, SO2_by_district.png, καθώς και ένα συγκεντρωτικό διάγραμμα συνολικής έκθεσης (Total_Pollutants_per_Capita.png).

Η πλήρης υλοποίηση είναι διαθέσιμη στο δημόσιο αποθετήριο MDAT-Scenario-2-Thessaloniki-Air-Quality-Population1, το οποίο ακολουθεί τη δομή που υπαγορεύει η μεθοδολογία του MDAT Pilot με διακριτούς φακέλους για την περιγραφή του σεναρίου (1_Scenario_Description_2), τα δεδομένα (2_data), τον κώδικα (3_code), τις εξόδους (4_Outputs) και τις πολιτικές μεταδεδομένων (5_Data_specification).

4.2.8.2 Αποτελέσματα και Οπτικοποιήσεις Σεναρίου 2

Η εκτέλεση της ροής επεξεργασίας απέδωσε ένα πλήρες αναλυτικό πλέγμα μετρήσεων για επτά ρύπους (CO, NO, NO₂, O₃, PM₁₀, PM_{2.5}, SO₂), σε πέντε δημοτικές κοινότητες της μητροπολιτικής περιοχής Θεσσαλονίκης οι οποίες αντιστοιχούν σε έξι σταθμούς μέτρησης (Εγνατίας, Λαγκαδά, Επταπυργίου, 25ης Μαρτίου, Μαλακοπής, Νέου Δημαρχείου), όπως αποτυπώνεται στο φύλλο Mapping του παραγόμενου αρχείου atmospheric_analysis_thessaloniki.xlsx. Η συνολική αναφορά συγκεντρώνεται στο φύλλο Συνολικοί Μέσοι & Ανά Κάτοικο, το οποίο περιλαμβάνει 39 εγγραφές μέσω συγκεντρώσεων με την αντίστοιχη ταξινόμηση συμμόρφωσης (dpv:ComplianceStatus).

Συνοπτική εικόνα συμμόρφωσης.

Από το σύνολο των μετρήσεων προκύπτουν πέντε περιπτώσεις υπέρβασης των θεσμοθετημένων ορίων της Ευρωπαϊκής Ένωσης, συγκεντρωμένες σχεδόν εξ ολοκλήρου στο 1η Δημοτική Κοινότητα (Σταθμός Εγνατίας), με μία επιπλέον υπέρβαση PM₁₀ στο 2η Δημοτική Κοινότητα (Σταθμός Λαγκαδά). Ο Πίνακας 4.3 συνοψίζει τις περιπτώσεις αυτές:

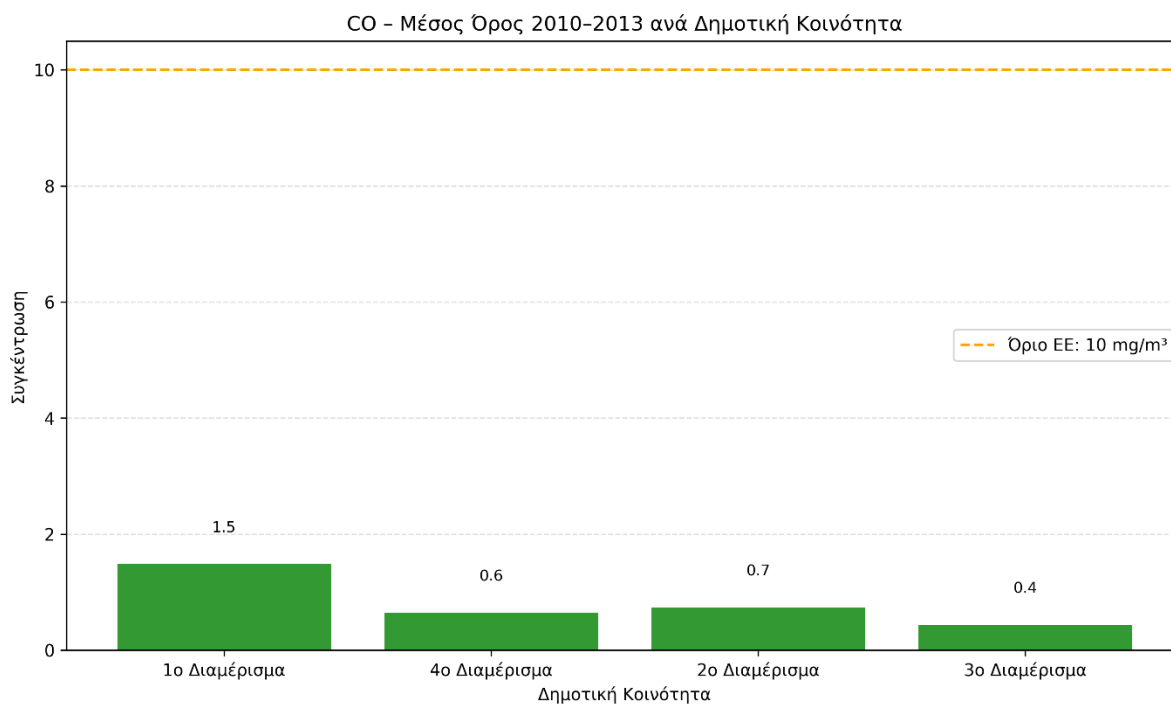
Ρύπος	Σταθμός	Δημοτική Κοινότητα	Μέση Τιμή 2010-2013	Όριο ΕΕ	Υπέρβαση
NO	Εγνατίας	1η Δημοτική Κοινότητα	108,9 $\mu\text{g}/\text{m}^3$	100 $\mu\text{g}/\text{m}^3$	+8,9%
NO ₂	Εγνατίας	1η Δημοτική Κοινότητα	85,6 $\mu\text{g}/\text{m}^3$	40 $\mu\text{g}/\text{m}^3$	+114,0%
PM ₁₀	Εγνατίας	1η Δημοτική Κοινότητα	49,4 $\mu\text{g}/\text{m}^3$	40 $\mu\text{g}/\text{m}^3$	+23,5%
PM _{2.5}	Εγνατίας	1η Δημοτική Κοινότητα	34,5 $\mu\text{g}/\text{m}^3$	25 $\mu\text{g}/\text{m}^3$	+38,0%
PM ₁₀	Λαγκαδά	2η Δημοτική Κοινότητα	47,3 $\mu\text{g}/\text{m}^3$	40 $\mu\text{g}/\text{m}^3$	+18,3%

Πίνακας 4.3: Καταγεγραμμένες υπερβάσεις των ευρωπαϊκών ορίων ποιότητας αέρα στον Δήμο Θεσσαλονίκης, περίοδος 2010-2013. Αποτέλεσμα της διεργασίας `mdat:AssessComplianceWithEULimits`.

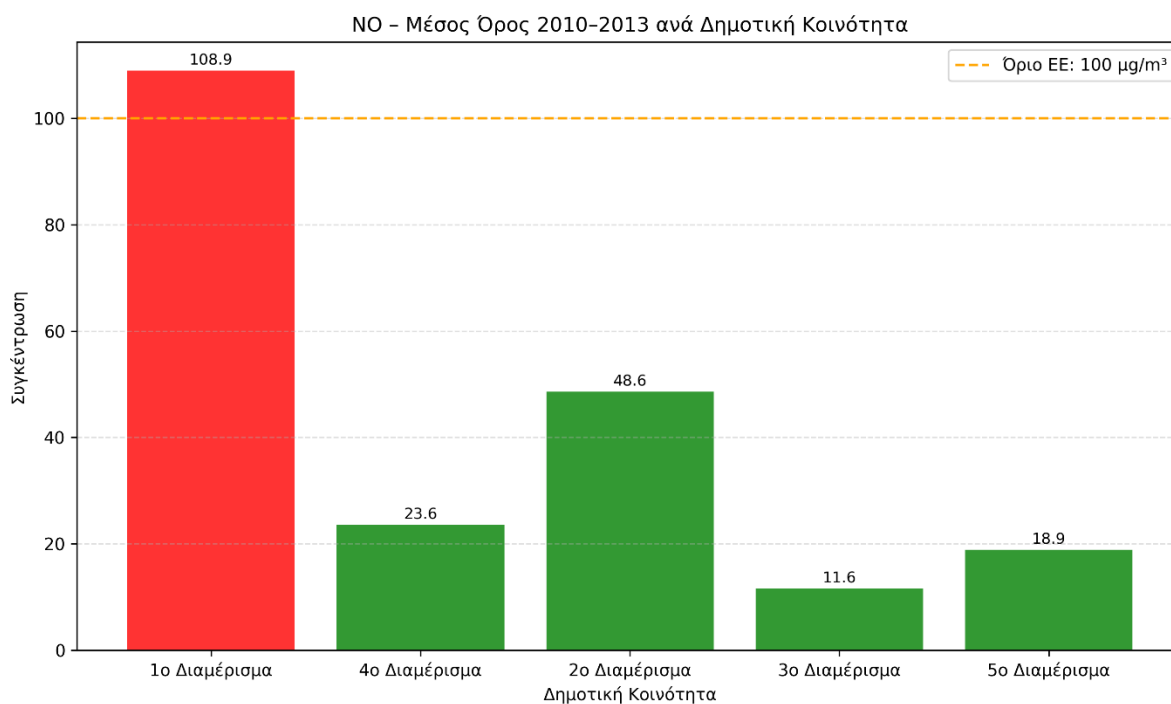
Η σαφής συγκέντρωση των υπερβάσεων στο 1η Δημοτική Κοινότητα (Σταθμός Εγνατίας) αποτελεί το πιο εμφανές εύρημα της ανάλυσης. Ιδιαίτερη ανησυχία προκαλεί η υπέρβαση του ορίου NO₂ κατά 114% (85,6 έναντι 40 $\mu\text{g}/\text{m}^3$), τιμή που υπερβαίνει υπερδιπλάσια το επιτρεπτό όριο και υποδεικνύει συστηματική επιβάρυνση από την κυκλοφορία οχημάτων στον κεντρικό οδικό άξονα της πόλης.

Ανάλυση κατά ρύπο:

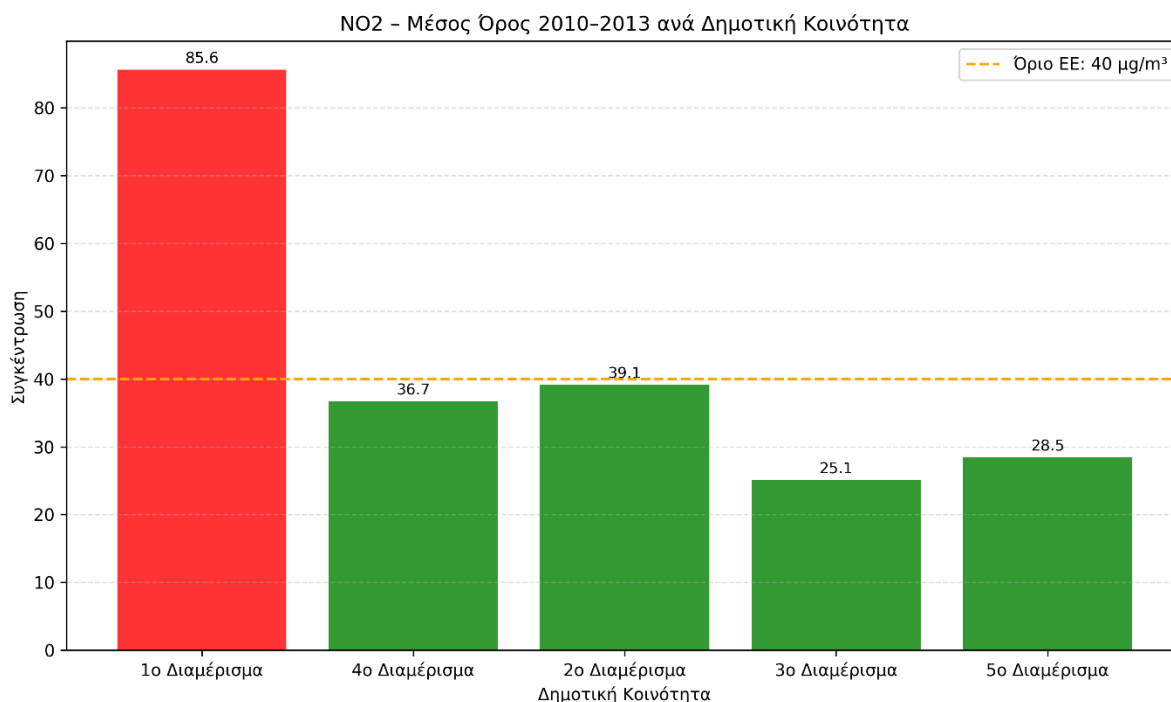
Τα Σχήματα 4.6 έως 4.12 παρουσιάζουν τη χωρική κατανομή κάθε ρύπου ανά δημοτική κοινότητα, με ομοιόμορφη χρωματική κωδικοποίηση (πράσινο για εντός ορίων, κόκκινο για υπέρβαση) που παράγεται αυτόματα από τη διεργασία `mdat:GenerateGraphsAndVisualSummaries`:



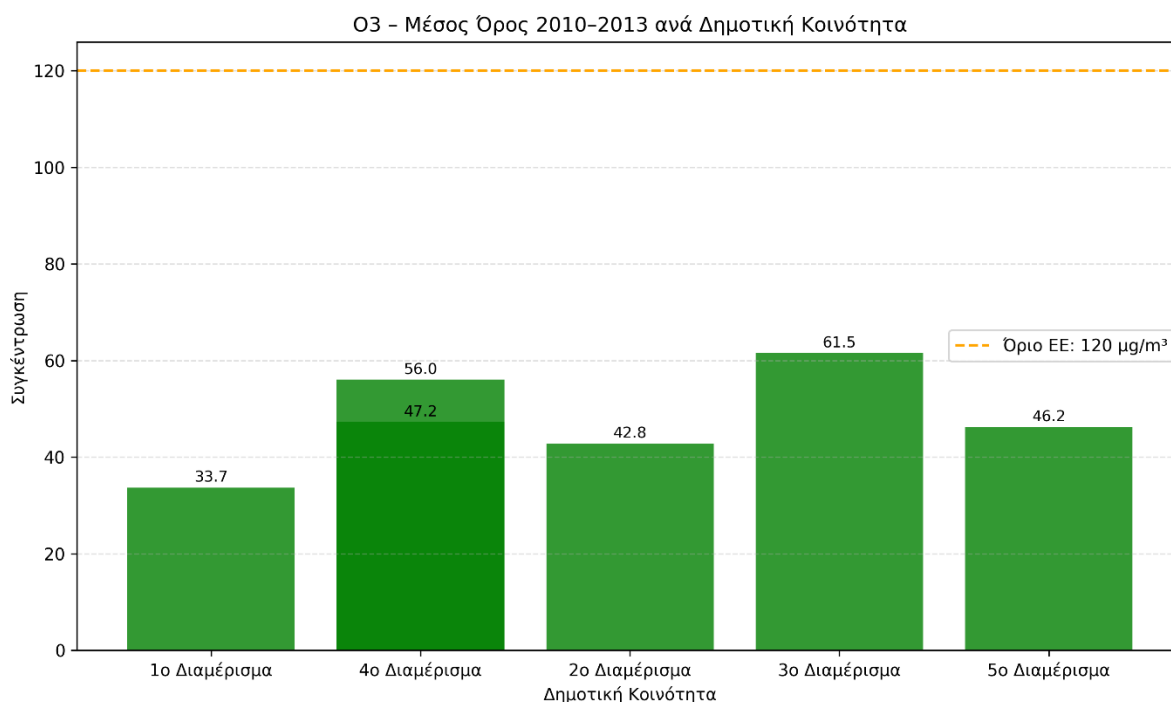
Σχήμα 4.6: Μέση συγκέντρωση μονοξειδίου του άνθρακα (CO) ανά δημοτική κοινότητα, 2010-2013. Όλες οι μετρήσεις παραμένουν σαφώς εντός του ορίου ΕΕ (10 mg/m³), με εύρος από 0,4 (3η Δημοτική Κοινότητα) έως 1,5 mg/m³ (1η Δημοτική Κοινότητα).



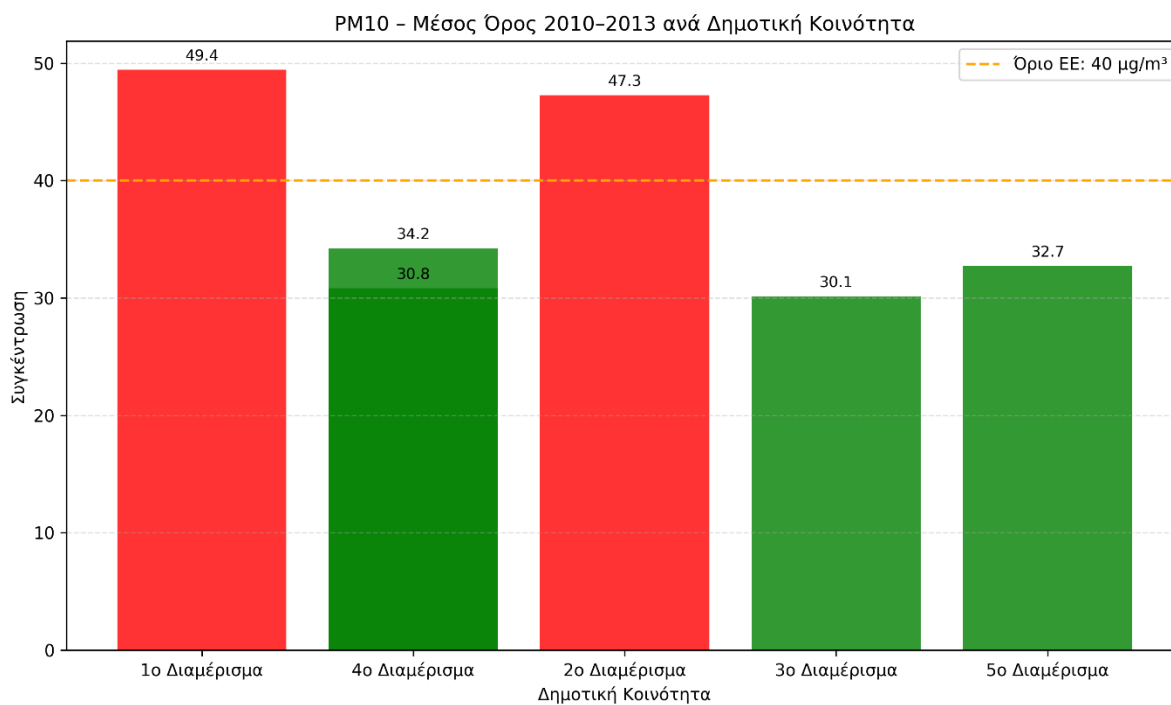
Σχήμα 4.7: Μέση συγκέντρωση μονοξειδίου του αζώτου (NO) ανά δημοτική κοινότητα. Καταγράφεται οριακή υπέρβαση του ορίου ΕΕ (100 µg/m³) στο 1η Δημοτική Κοινότητα (108,9 µg/m³), ενώ οι λοιπές κοινότητες παραμένουν εντός ορίων με τιμές 11,6-48,6 µg/m³.



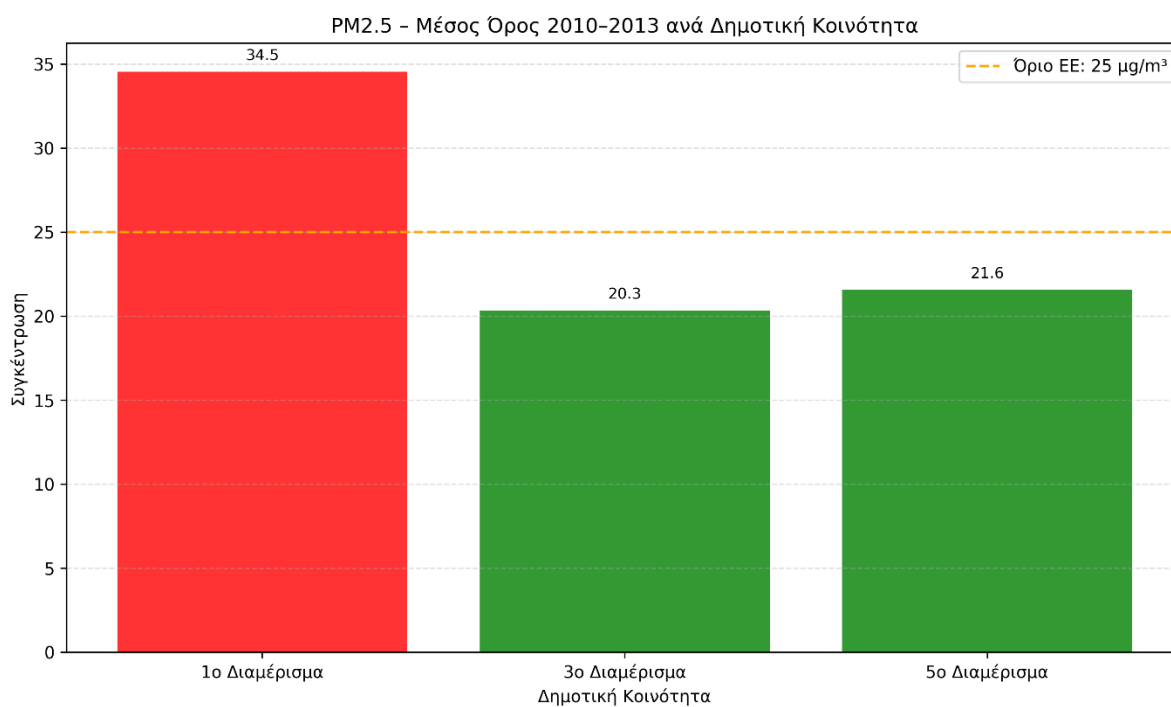
Σχήμα 4.8: Μέση συγκέντρωση διοξειδίου του αζώτου (NO₂) ανά δημοτική κοινότητα. Η κρισιμότερη υπέρβαση του σεναρίου: το 1η Δημοτική Κοινότητα καταγράφει 85,6 µg/m³ έναντι ορίου 40 µg/m³ (υπέρβαση 114%). Το 2η Δημοτική Κοινότητα (39,1 µg/m³) προσεγγίζει οριακά το όριο.



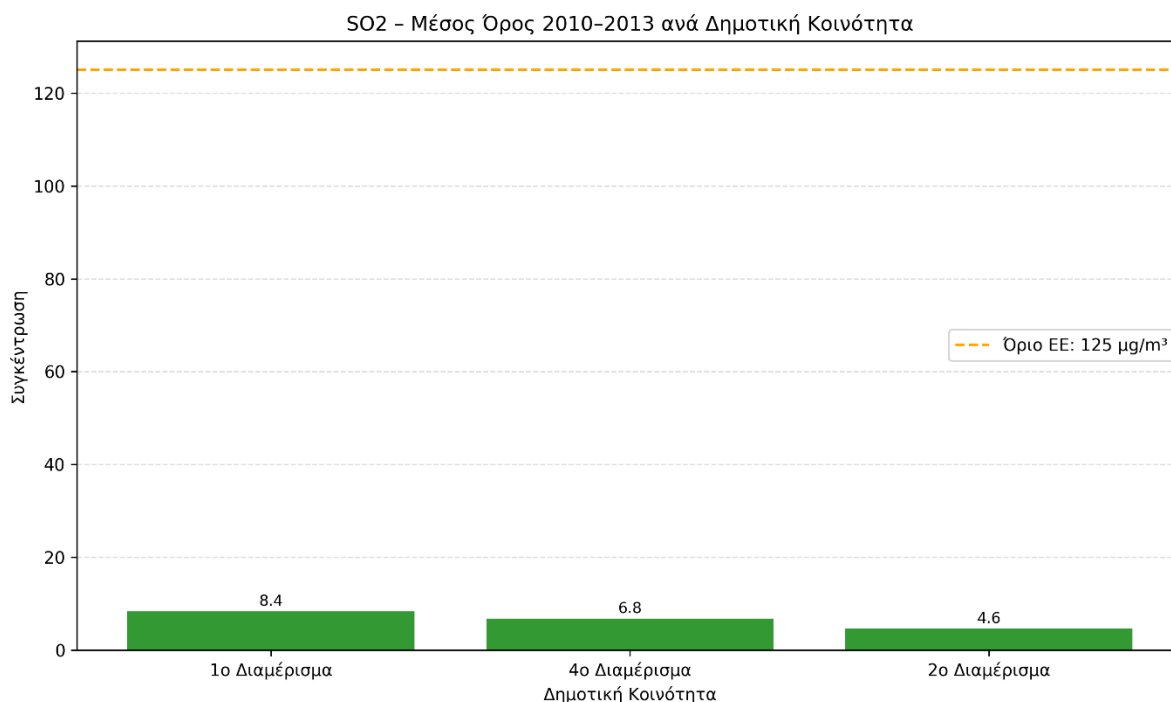
Σχήμα 4.9: Μέση συγκέντρωση όζοντος (O₃) ανά δημοτική κοινότητα. Όλες οι τιμές (33,7-61,5 µg/m³) παραμένουν εντός του ορίου ΕΕ (120 µg/m³). Αξιοσημείωτο είναι το αντίστροφο μοτίβο σε σχέση με τα NO/NO₂: η μέγιστη τιμή O₃ εμφανίζεται στο 3η Δημοτική Κοινότητα (61,5 µg/m³), περιοχή με χαμηλή κυκλοφοριακή φόρτιση-φαινόμενο συμβατό με τη γνωστή φωτοχημική συμπεριφορά του O₃.



Σχήμα 4.10: Μέση συγκέντρωση αιωρούμενων σωματιδίων PM_{10} ανά δημοτική κοινότητα. Καταγράφονται δύο υπερβάσεις του ορίου ΕΕ (40 µg/m³): στο 1η Δημοτική Κοινότητα (49,4 µg/m³) και στο 2η Δημοτική Κοινότητα (47,3 µg/m³). Οι υπόλοιπες κοινότητες παραμένουν εντός ορίων με μικρές διαφορές μεταξύ τους (30,1-32,7 µg/m³).



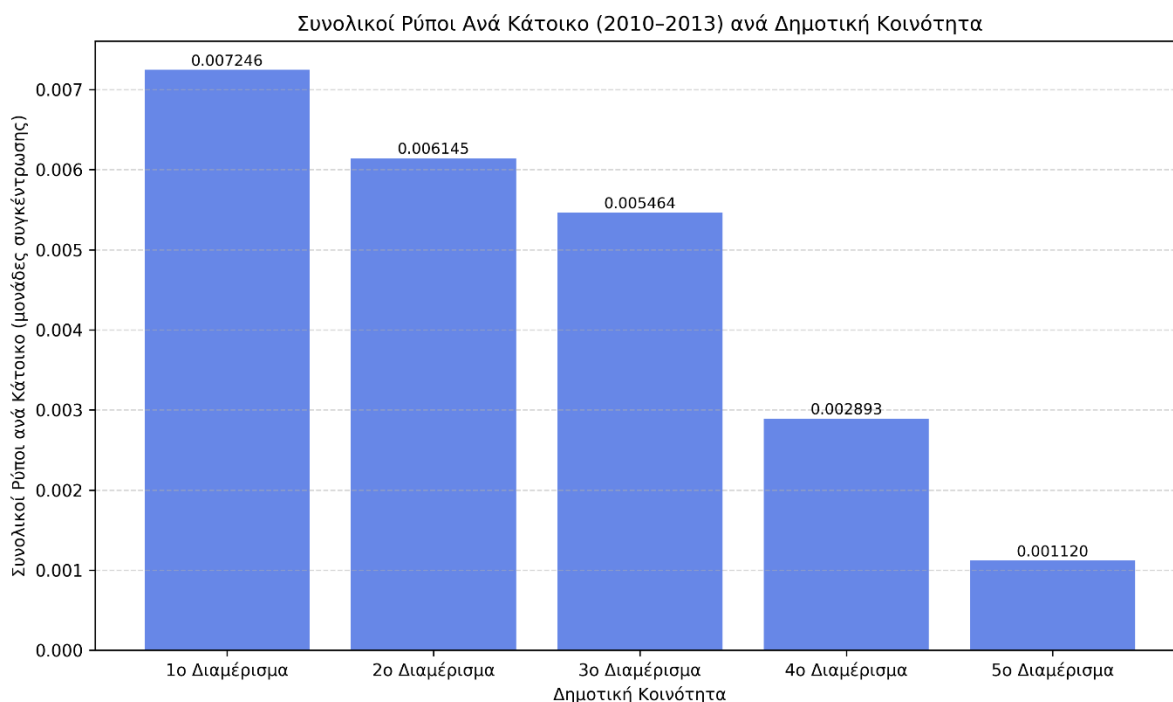
Σχήμα 4.11: Μέση συγκέντρωση αιωρούμενων σωματιδίων $PM_{2.5}$ ανά δημοτική κοινότητα. Καταγράφεται υπέρβαση του ορίου ΕΕ (25 µg/m³) στο 1η Δημοτική Κοινότητα (34,5 µg/m³, υπέρβαση 38%), ενώ το 3ο και 5η Δημοτική Κοινότητα παραμένουν εντός ορίων (20,3 και 21,6 µg/m³ αντίστοιχα). Δεδομένα $PM_{2.5}$ δεν είναι διαθέσιμα για όλες τις κοινότητες, στοιχείο που υποδεικνύει μερική κάλυψη του δικτύου παρακολούθησης.



Σχήμα 4.12: Μέση συγκέντρωση διοξειδίου του θείου (SO₂) ανά δημοτική κοινότητα. Όλες οι μετρήσεις παραμένουν δραματικά κάτω από το όριο ΕΕ (125 μg/m³), με μέγιστη τιμή 8,4 μg/m³ στο 1η Δημοτική Κοινότητα. Το εύρημα αντικατοπτρίζει τη γενικότερη μείωση του SO₂ σε αστικά περιβάλλοντα μετά τον περιορισμό της χρήσης πετρελαίου θέρμανσης υψηλής περιεκτικότητας σε θείο.

Συνθετικός δείκτης έκθεσης ανά κάτοικο

Πέρα από την ανά-ρύπο ανάλυση, η διεργασία `mdat:ComputePollutantPerCapitaRatios` υπολόγισε έναν σύνθετο δείκτη συνολικής έκθεσης ανά κάτοικο (`mdat:AirPollutionIndex`), ο οποίος αποτυπώνει αθροιστικά τη ρυπαντική επιβάρυνση κανονικοποιημένη ως προς τον μόνιμο πληθυσμό κάθε δημοτικής κοινότητας.



Σχήμα 4.13: Συνθετικός δείκτης συνολικής έκθεσης ανά κάτοικο (*mdat:AirPollutionIndex*) ανά δημοτική κοινότητα, 2010-2013. Αποτέλεσμα της διεργασίας *mdat:ComputePollutantPerCapitaRatios*.

4.2.9 Συμπεράσματα του Σεναρίου

Το δεύτερο σενάριο επεκτείνει τη συνεισφορά του πρώτου σε δύο κρίσιμες διαστάσεις. Πρώτον, υποστηρίζει τη δυνατότητα του πλαισίου να ενοποιεί δεδομένα διαφορετικού καθεστώτος πρόσβασης (ανοικτά και περιορισμένα) στο πλαίσιο μιας ενιαίας, σημασιολογικά τεκμηριωμένης ροής. Δεύτερον, αναδεικνύει εμπράκτως τον ρόλο του UPCAST Negotiation Plugin ως μηχανισμού επίσημης διαπραγμάτευσης και του consumer Dashboard ως εκτελεστή της αυτόματης επανασχεδίασης workflows, στοιχεία που υποδεικνύει ότι η αρχή του Compliance-by-Design από θεωρητική επιδίωξη μπορεί να λειτουργήσει ως πραγματικός μηχανισμός συμμόρφωσης.

Πέραν της μεθοδολογικής συνεισφοράς, η ανάλυση ανέδειξε και ευρήματα ουσιαστικής αξίας: ενώ ορισμένοι ρύποι (CO, O₃, SO₂) παραμένουν εντός των ευρωπαϊκών ορίων σε όλες τις δημοτικές κοινότητες, καταγράφονται σαφείς υπερβάσεις σε NO, NO₂, PM₁₀ και PM_{2.5}, με ιδιαίτερη συγκέντρωση στην 1η Δημοτική Κοινότητα (Σταθμός Εγνατίας). Η σημαντική διακύμανση των μετρήσεων μεταξύ των δημοτικών κοινοτήτων υποδεικνύει ότι η συγκεντρωτική αξιολόγηση σε επίπεδο δήμου ενδέχεται να αποκρύπτει τοπικές ανισότητες έκθεσης. Η σημασιολογική διασύνδεση των μετρήσεων ρύπων με τον αντίστοιχο πληθυσμιακό δείκτη επιτρέπει την ανάδειξη τέτοιων ασυμμετριών με τρόπο ιχνηλάσιμο και επαναχρησιμοποιήσιμο.

Τέλος, αξίζει να σημειωθούν δύο περιορισμοί που ορίζουν φυσικές προεκτάσεις της παρούσας μελέτης. Αφενός, τα δεδομένα ρύπανσης αφορούν την περίοδο 2010-2013 και η επικαιροποίησή τους θα επέτρεπε διαχρονική σύγκριση και αξιολόγηση της επίδρασης τυχόν περιβαλλοντικών παρεμβάσεων. Αφετέρου, η ανάλυση περιορίζεται σε πέντε δημοτικές κοινότητες λόγω της κάλυψης του δικτύου μέτρησης, ενώ η μελλοντική ενσωμάτωση δεδομένων δορυφορικής παρακολούθησης θα μπορούσε να επεκτείνει σημαντικά τη χωρική ανάλυση. Παρά τους περιορισμούς αυτούς, το σενάριο αποτελεί ολοκληρωμένη απόδειξη της προστιθέμενης αξίας που προσφέρει η σημασιολογικά τεκμηριωμένη, policy-aware

προσέγγιση σε σύνθετα οικοσυστήματα Data Spaces όπου ανοικτά και περιορισμένα δεδομένα συνυπάρχουν.

4.3 Σενάριο 3: Ενεργειακές και Περιβαλλοντικές Μετρήσεις

4.3.1 Πλαίσιο και Ερευνητικός Στόχος

Το τρίτο σενάριο μελέτης εστιάζει στην ποσοτικοποίηση του ενεργειακού αποτυπώματος και των αντίστοιχων έμμεσων εκπομπών διοξειδίου του άνθρακα (CO₂) στη Θεσσαλονίκη, σε ορίζοντα δύο δεκαετιών (1993-2012). Η μελέτη αξιοποιεί στατιστικά δεδομένα κατανάλωσης ηλεκτρικής ενέργειας ανά κατηγορία χρήσης και τα μετασχηματίζει σε δείκτες περιβαλλοντικής επίπτωσης μέσω εφαρμογής χρονομεταβλητού συντελεστή εκπομπών, ο οποίος αντικατοπτρίζει την πραγματική ιστορική ένταση άνθρακα του ελληνικού δικτύου.

Σε αντιδιαστολή με τα δύο προηγούμενα σενάρια, το παρόν παρουσιάζει ορισμένες ιδιαίτερες σχεδιαστικές επιλογές που αξίζει να επισημανθούν. Πρώτον, λειτουργεί αποκλειστικά σε μη-προσωπικά, δημόσια διαθέσιμα δεδομένα (dpv:NonPersonalData), χωρίς να απαιτείται διαπραγμάτευση πρόσβασης ή αυτόματη επανασχεδίαση του workflow, στοιχείο που το καθιστά παράδειγμα «ομαλής» ροής εργασίας σε ένα Data Space. Δεύτερον, εισάγει τη διάσταση της διαχρονικής ανάλυσης (1993-2012), επιτρέποντας την παρακολούθηση εξελίξεων και τάσεων που δεν ήταν δυνατές στα προηγούμενα σενάρια συγχρονικού χαρακτήρα. Τρίτον, αναδεικνύει τη μεθοδολογική σημασία της έμμεσης παραγωγής δεικτών: οι εκπομπές CO₂ δεν μετριοούνται απευθείας, αλλά υπολογίζονται μέσω εξωτερικού επιστημονικού παραμέτρου (συντελεστής εκπομπών) που μεταβάλλεται διαχρονικά, αντικατοπτρίζοντας την εξέλιξη του ενεργειακού μείγματος.

Ο κεντρικός ερευνητικός στόχος συνίσταται στην απάντηση τριών διασυνδεδεμένων ερωτημάτων: (α) ποια είναι η διαχρονική εξέλιξη της ηλεκτρικής κατανάλωσης ανά τομέα δραστηριότητας στη Θεσσαλονίκη, (β) πώς μεταφράζεται αυτή η κατανάλωση σε ποσοτικοποιημένο αποτύπωμα άνθρακα, λαμβάνοντας υπόψη τη σταδιακή απολιγνιτοποίηση του ελληνικού δικτύου, και (γ) η δημιουργία ενός επαναχρησιμοποιήσιμου workflow που μπορεί να εκτελεστεί ξανά με την ύπαρξη νέων παραλλαγών δεδομένων.

4.3.2 Πηγές Δεδομένων

Στο σενάριο αξιοποιούνται δύο σύνολα δεδομένων εισόδου και παράγεται ένα τελικό σύνολο εξόδου:

- D1 - Energy Consumption Thessaloniki (1993-2012): Στατιστικά δεδομένα ηλεκτρικής κατανάλωσης (kWh) ανά γεωγραφική περιοχή και κατηγορία χρήσης για την περίοδο 1993-2012, διαθέσιμα μέσω της πλατφόρμας Thessaloniki Data Space (TDS) - OKF Greece. Σημασιολογικά κατηγοριοποιείται ως dpv:EnvironmentalData / dpv:NonPersonalData και αντιστοιχίζεται στην έννοια mdat:EnergyDataset. Περιλαμβάνει έξι κατηγορίες χρήσης: Οικιακή (Domestic), Εμπορική (Commercial), Βιομηχανική (Industrial), Γεωργική (Agricultural), Δημόσιες & Δημοτικές Αρχές (Public) και Φωτισμός Οδών (Lighting).
- D2 - Greek Grid Carbon Intensity Time-Series (1993-2025): Χρονοσειρά συντελεστών έντασης άνθρακα της ελληνικής ηλεκτροπαραγωγής, βάσει της επίσημης βάσης δεδομένων της Ember Climate (Yearly Electricity Data, 2025), διαθέσιμη μέσω του Our World in Data2. Σημασιολογικά αποτελεί dpv:EnvironmentalData / dpv:NonPersonalData και αντιστοιχίζεται στη νέα έννοια mdat:GridCarbonIntensitySeries. Οι τιμές υπολογίζονται με μεθοδολογία lifecycle

emissions, συμπεριλαμβανομένων των upstream εκπομπών μεθανίου, αλυσίδας εφοδιασμού και κατασκευής σταθμών, εκφρασμένες σε ισοδύναμα CO₂ σε ορίζοντα 100 ετών.

- D3 - Derived Energy and Emissions Output: Παραγόμενο σύνολο που περιλαμβάνει το αρχείο thess_energy_analysis.xlsx (τρία φύλλα: Analysis, EmissionFactors, Metadata) και τα συνοδευτικά γραφήματα PNG. Χαρακτηρίζεται ως dpv:DerivedData και αντιστοιχίζεται στην έννοια mdat:CO2EmissionDataset.

Συντελεστής Εκπομπών CO₂. Ιδιαίτερη μεθοδολογική πρόκληση του σεναρίου αποτελεί η επιλογή κατάλληλου συντελεστή εκπομπών. Σύμφωνα με τα επίσημα δεδομένα της Ember, ο συντελεστής έντασης άνθρακα της ελληνικής ηλεκτροπαραγωγής κυμάνθηκε μεταξύ 704,06 και 877,02 g CO₂/kWh κατά την περίοδο 1993–2012, με μέσο όρο ~794 g CO₂/kWh και πιο πρόσφατη μετρημένη τιμή τα 315,09 g CO₂/kWh για το 2025. Η σημαντική αυτή μείωση αντικατοπτρίζει τη σταδιακή απολιγνιτοποίηση του ελληνικού δικτύου μέσω της εισόδου του φυσικού αερίου και της ανάπτυξης των Ανανεώσιμων Πηγών Ενέργειας.

Για την ολοκληρωμένη αποτύπωση του ιστορικού περιβαλλοντικού αποτυπώματος, το σενάριο εφαρμόζει διπλή εκτίμηση εκπομπών: μια αναφορά με χρήση της πιο πρόσφατης διαθέσιμης τιμής (mdat:emissionFactor_reference = 0,31509 kg/kWh, έτος αναφοράς 2025) και μια ιστορικά συμβατή με χρήση διαφορετικού συντελεστή για κάθε έτος (mdat:emissionFactor_historical), που αποτυπώνει την πραγματική ένταση άνθρακα τη στιγμή της κατανάλωσης. Η σημασιολογική αποτύπωση της παραμέτρου ως χρονομεταβλητής (mdat:TimeVaryingParameter) εξασφαλίζει την πλήρη ιχνηλασιμότητα κάθε ενδιαμέσου υπολογισμού. Σημειώνεται ότι αμφότεροι οι συντελεστές προέρχονται από την ίδια πηγή (Ember) και την ίδια μεθοδολογία (lifecycle emissions), εξασφαλίζοντας μεθοδολογική συνέπεια στις διαχρονικές συγκρίσεις.

4.3.3 Ρόλοι Οντολογίας

Σύμφωνα με το DPV 2.2, στο σενάριο αναγνωρίζονται τρεις διακριτοί ρόλοι:

- Data Provider (dpv:DataSource, dpv:DataController): Παρέχει τα επίσημα στατιστικά δεδομένα ηλεκτρικής κατανάλωσης μέσω της πλατφόρμας TDS - OKFN Greece, καθώς και τη χρονοσειρά συντελεστών έντασης άνθρακα μέσω της Ember Climate, υπό όρους ανοικτής διάθεσης.
- Data Analyst (dpv:DataProcessor): Εκτελεί την τεχνική επεξεργασία των δεδομένων μέσω του Python script, υλοποιεί τον υπολογισμό των εκπομπών CO₂ και παράγει τις οπτικοποιήσεις.
- Researcher / Policy User (dpv:DataUser, dpv:Recipient): Ερμηνεύει τα παραγόμενα αποτελέσματα και τα μεταφράζει σε εισηγήσεις για ενεργειακή πολιτική και πολιτικές βιωσιμότητας σε επίπεδο τοπικής αυτοδιοίκησης.

Όπως και στο πρώτο σενάριο, η απουσία dpv:RestrictedData καθιστά τη ροή απαλλαγμένη από αυτόματους κόμβους συμμόρφωσης (odrl:delete, dpv:Store), διατηρώντας την απλότητα ενός open-data workflow.

4.3.4 Σημασιολογική Αντιστοίχιση Διεργασιών

Η ροή επεξεργασίας απαρτίζεται από τρία διακριτά βήματα, καθένα εκ των οποίων μοντελοποιείται ως dpv:Process και αντιστοιχίζεται σε κλάσεις DPV, actions ODRL και εξειδικευμένες έννοιες του namespace mdat:. Ο Πίνακας 4.5 παρουσιάζει την πλήρη αντιστοίχιση:

Βήμα	Κόμβος	DPV	ODRL	MDAT Concept	Ρόλος
1	CleanAndNormalise	dpv:Transform	odrl:modify	mdat:SelectSubset	Data Analyst
2	ComputeCO2Emissions	dpv:Derive	odrl:derive	mdat:ComputeCO2Emissions, mdat:CO2EmissionDataset	Data Analyst
3	GenerateEmissionVisuals	dpv:Visualise, dpv:Analyse	odrl:display, reproduce	mdat:EnergyEmissionVisualisation	Researcher

Πίνακας 4.5: Σενάριο 3 - Αποσύνθεση διεργασιών και σημασιολογική αντιστοίχιση.

Η συντομία του πίνακα έναντι των αντίστοιχων στα Σενάρια 1 και 2 αποτυπώνει τη μειωμένη πολυπλοκότητα της ροής, η οποία αποτελεί διδακτικό παράδειγμα *minimal yet complete policy-aware workflow*. Ιδιαίτερη σημασία έχει η προσθήκη του *dpv:Combine* στο βήμα 2, καθώς ο υπολογισμός εκπομπών συνδυάζει δύο διακριτά datasets: το dataset κατανάλωσης (D1) και τη χρονοσειρά συντελεστών (D2), με την αντιστοίχιση να γίνεται μέσω του πεδίου Year.

4.3.5 Ροή Επεξεργασίας Δεδομένων

Η αρχιτεκτονική του workflow, όπως αποτυπώνεται στο Σχήμα 4.14, ακολουθεί ένα γραμμικό σειριακό μοτίβο εκτέλεσης χωρίς διακλαδώσεις:

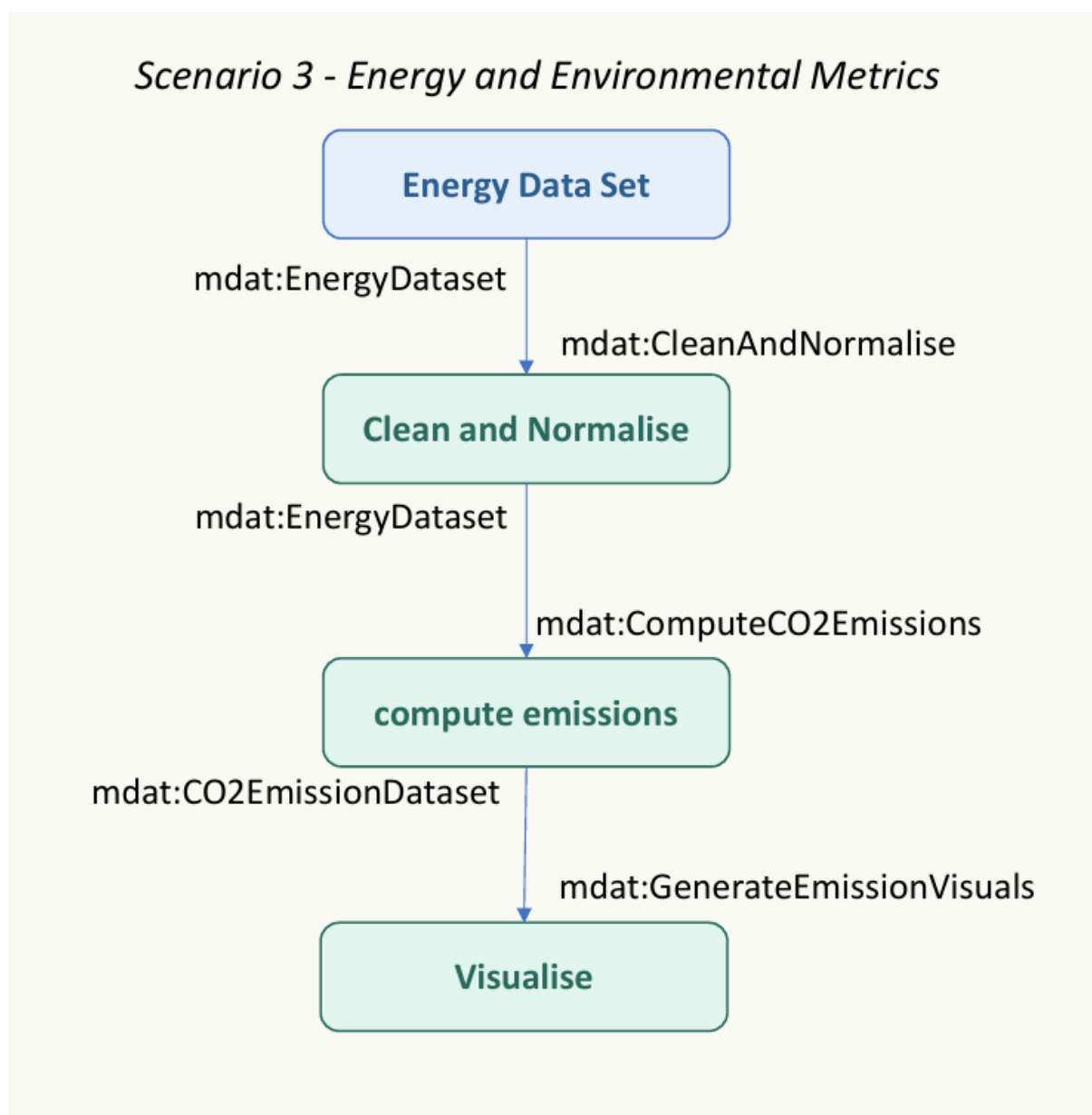
CleanAndNormalise: Δέχεται ως είσοδο το dataset *mdat:EnergyDataset* σε μορφή Excel και εκτελεί τέσσερις επιμέρους εργασίες: μετονομασία και κανονικοποίηση του σχήματος των στηλών, φιλτράρισμα του υποσυνόλου που αφορά τη Θεσσαλονίκη, μετατροπή των αριθμητικών πεδίων σε τύπο δεδομένων float (με αντιμετώπιση του ελληνικού δεκαδικού διαχωριστικού), και εξυγίανση των χρονικών πεδίων. Η αντιστοίχιση με *dpv:Transform* και *odrl:modify* αποτυπώνει τον αμιγώς μετασχηματιστικό χαρακτήρα του βήματος.

ComputeCO2Emissions: Αποτελεί το αναλυτικό επίκεντρο του σεναρίου. Πραγματοποιείται διπλός υπολογισμός εκπομπών: ένας με τη χρονομεταβλητή ιστορική τιμή του συντελεστή για κάθε έτος, και ένας με την τιμή αναφοράς του 2025 (315,09 g CO₂/kWh). Ο τύπος εφαρμόζεται ξεχωριστά σε καθεμία από τις έξι κατηγορίες χρήσης και για κάθε έτος της περιόδου 1993-2012:

$$\text{CO}_2(\text{kg})_{y,c} = \text{Consumption}_{y,c}(\text{kWh}) \times \text{EF}_y(\text{g/kWh})/1000$$

όπου y το έτος, c η κατηγορία χρήσης, και EF_y ο συντελεστής εκπομπών για το έτος y . Το αποτέλεσμα συγκροτεί την οντότητα *mdat:CO2EmissionDataset*. Ο διπλός χαρακτηρισμός *dpv:Derive* + *dpv:Combine* αποτυπώνει τόσο τη συνδυαστική (κατανάλωση × συντελεστής) όσο και την παραγωγική διάσταση της ενέργειας.

GenerateEmissionVisuals: Παράγει πέντε διακριτές οπτικοποιήσεις (παρουσιάζονται στην ενότητα 4.3.6.2): την κατανάλωση ανά κατηγορία, τις εκπομπές με ιστορικό συντελεστή, την εξέλιξη του συντελεστή κατά την περίοδο μελέτης, την επεκταμένη εξέλιξη του μέχρι το 2025, και τη συγκριτική αποτύπωση των δύο σεναρίων υπολογισμού. Ο διπλός χαρακτηρισμός *dpv:Visualise* + *dpv:Analyse* αντικατοπτρίζει τόσο την οπτική αναπαράσταση όσο και την αναλυτική διάσταση της απεικόνισης τάσεων.



Σχήμα 4.14: DPW - Σενάριο 3: Ενεργειακές και Περιβαλλοντικές Μετρήσεις

4.3.6 Δομή Παραδοτέων και Αποτελέσματα Εκτέλεσης

4.3.6.1 Δομή Παραδοτέων κατά την εκτέλεση του workflow

Τα αναμενόμενα παραδοτέα που θα λαμβάνονται σε κάθε εκτέλεση του workflow θα περιλαμβάνουν τους παρακάτω τύπους δεδομένων:

- Το αρχείο *thess_energy_analysis.xlsx* με τρία φύλλα:
 - *Analysis*: Αναλυτικός πίνακας 20 εγγραφών με κατανάλωση και εκπομπές CO₂ ανά κατηγορία και έτος, σύμφωνα και με τους δύο συντελεστές (dpv:DerivedData, mdat:CO2EmissionDataset).
 - *EmissionFactors*: Πίνακας συντελεστών εκπομπών ανά έτος με αναφορά πηγής.
 - *Metadata*: Πλήρης σημασιολογική τεκμηρίωση του παραδοτέου (DPV/ODRL/MDAT mapping, μεθοδολογία, πηγές).
- Πέντε γραφήματα σε μορφή PNG (dpv:VisualisationData, mdat:EnergyEmissionVisualisation):

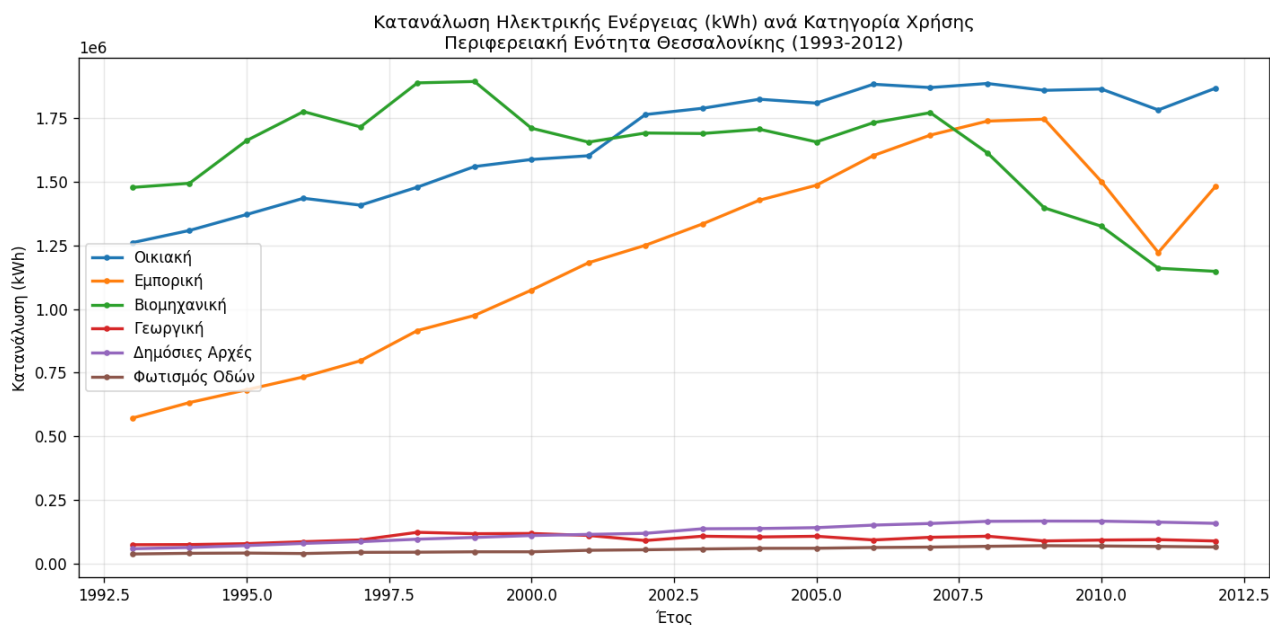
- Energy_consumption_per_category.png - διαχρονική εξέλιξη κατανάλωσης.
- co2_emissions_per_category.png - διαχρονική εξέλιξη εκπομπών CO₂ ανά κατηγορία.
- emission_factor_evolution.png - εξέλιξη του συντελεστή 1993-2012.
- emission_factor_extended_1993_2025.png - επεκταμένη εξέλιξη του συντελεστή 1993-2025.
- co2_total_current_vs_historical.png - συγκριτική αποτύπωση των δύο σεναρίων.

Η πλήρης υλοποίηση είναι διαθέσιμη στο δημόσιο αποθετήριο SCENARIO-3, το οποίο ακολουθεί τη δομή του MDAT Pilot.

4.3.6.2 Αποτελέσματα και Οπτικοποιήσεις Σεναρίου 3

Η εκτέλεση της ροής απέδωσε ένα ολοκληρωμένο διαχρονικό προφίλ της ηλεκτρικής κατανάλωσης και των αντίστοιχων εκπομπών CO₂ στη Θεσσαλονίκη για είκοσι έτη. Τα ευρήματα παρουσιάζονται σε πέντε συμπληρωματικά γραφήματα.

Διαχρονική εξέλιξη ηλεκτρικής κατανάλωσης. Το Σχήμα 4.15 αποτυπώνει την κατανάλωση ηλεκτρικής ενέργειας ανά κατηγορία χρήσης. Τρεις κατηγορίες κυριαρχούν σαφώς στο ενεργειακό προφίλ της πόλης (Οικιακή, Εμπορική, Βιομηχανική), ενώ οι υπόλοιπες τρεις (Γεωργική, Δημόσιες Αρχές, Φωτισμός Οδών) παρουσιάζουν σταθερά χαμηλές τιμές καθ' όλη την περίοδο μελέτης.

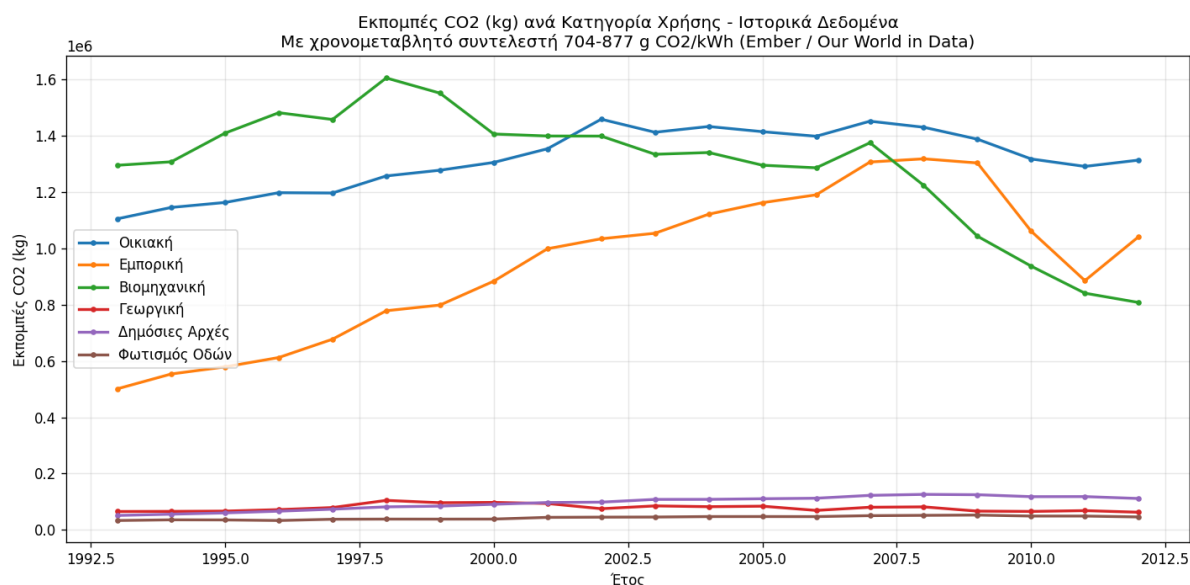


Σχήμα 4.15: Διαχρονική εξέλιξη της ηλεκτρικής κατανάλωσης (kWh) ανά κατηγορία χρήσης στη Θεσσαλονίκη, 1993-2012.

Από την ανάλυση του διαγράμματος προκύπτουν τέσσερα ποιοτικά ευρήματα: (1) Σταθερή πρωτοκαθεδρία της Οικιακής χρήσης, που αυξάνεται από ~1,26 σε ~1,87 εκατ. kWh (αύξηση ~48%). (2) Ραγδαία ανάπτυξη της Εμπορικής χρήσης, με σχεδόν τριπλασιασμό από ~0,57 σε ~1,48 εκατ. kWh, στοιχείο συμβατό με την εκρηκτική ανάπτυξη του τριτογενούς τομέα. (3) Πτωτική τάση της Βιομηχανικής χρήσης μετά το 2008, με υποχώρηση σε ~1,15 εκατ. kWh το 2012, χρονικά συμπίπτουσα με την έναρξη της οικονομικής κρίσης. (4) Αντιστροφή ιεραρχίας περί το 2002-2006, όπου η Οικιακή κατανάλωση ξεπερνά σταθερά τη Βιομηχανική.

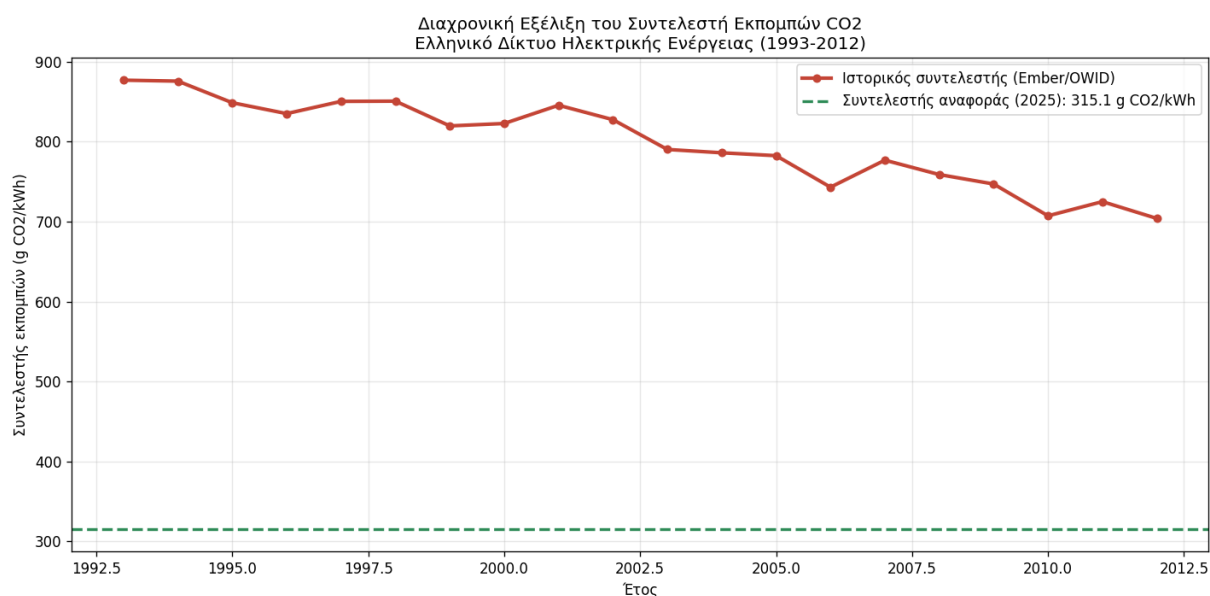
Διαχρονική εξέλιξη εκπομπών CO₂. Το Σχήμα 4.16 παρουσιάζει τις αντίστοιχες εκπομπές CO₂ με χρήση του χρονομεταβλητού ιστορικού συντελεστή. Σε αντιδιαστολή με ένα γραμμικό

μετασχηματισμό, το γράφημα αποκτά διαφορετική μορφή από το Σχήμα 4.15: οι εκπομπές κορυφώνονται γύρω στο 2007 (~4,4 εκατ. kg CO₂ συνολικά) και έκτοτε εμφανίζουν πτωτική τάση τόσο λόγω μείωσης της κατανάλωσης όσο και λόγω σταδιακής μείωσης του συντελεστή έντασης άνθρακα.



Σχήμα 4.16: Διαχρονική εξέλιξη των έμμεσων εκπομπών CO₂ (kg) ανά κατηγορία χρήσης στη Θεσσαλονίκη, 1993-2012. Υπολογισμός με χρονομεταβλητό συντελεστή 704-877 g CO₂/kWh (Ember/OWID).

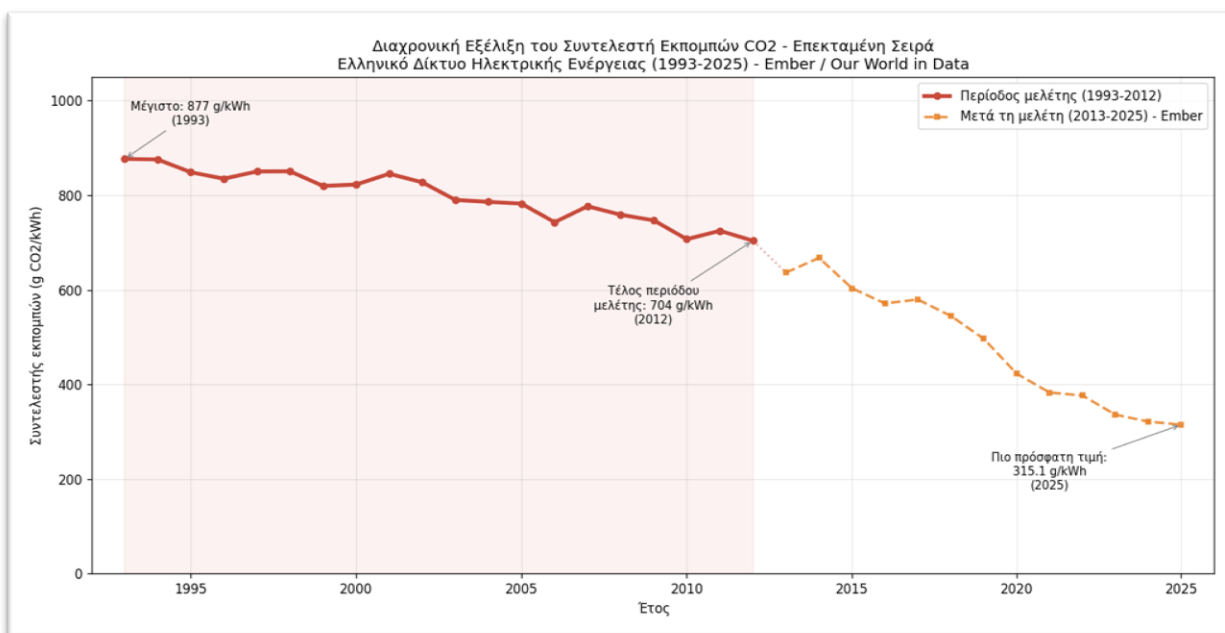
Εξέλιξη του συντελεστή εκπομπών κατά την περίοδο μελέτης. Το Σχήμα 4.17 παρουσιάζει την εξέλιξη του ίδιου του συντελεστή έντασης άνθρακα του ελληνικού δικτύου κατά την περίοδο 1993-2012, με ταυτόχρονη οπτική παράθεση της τιμής αναφοράς του 2025. Παρατηρείται μια μη μονοτονική πτωτική τάση: από 877 g/kWh το 1993 σε 704 g/kWh το 2012, με αναποδιές το 2001 (845,6) και 2007 (776,9), που υποδεικνύουν περιόδους αυξημένης χρήσης λιγνίτη. Η συνολική μείωση κατά την περίοδο μελέτης ανέρχεται σε ~20%.



Σχήμα 4.17: Διαχρονική εξέλιξη του συντελεστή εκπομπών CO₂ για το ελληνικό δίκτυο, 1993-2012. Πηγή: Ember

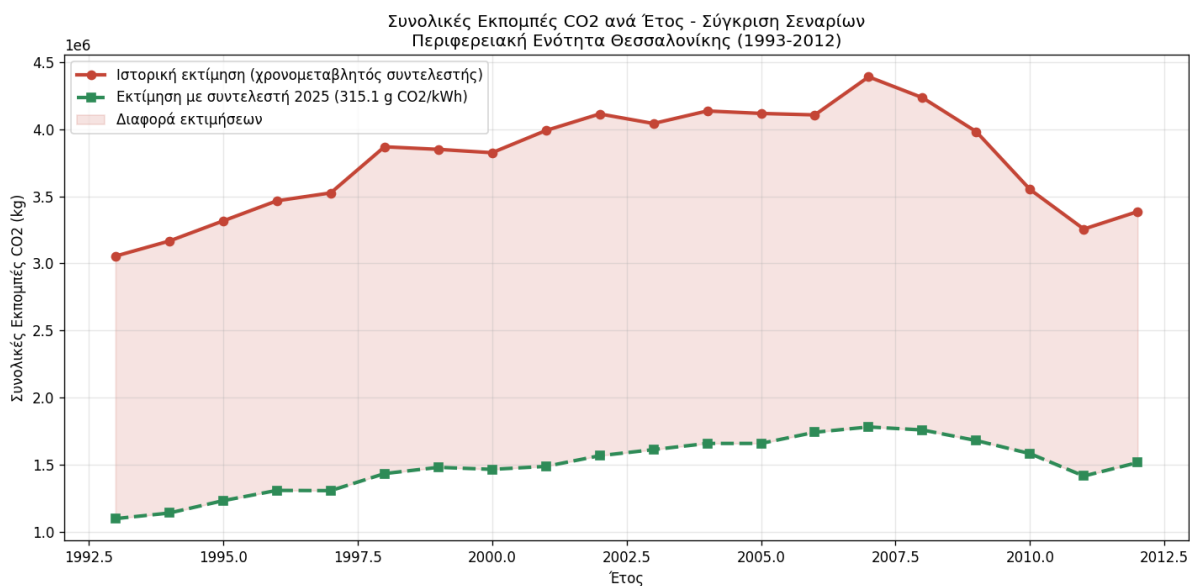
Yearly Electricity Data (2025) - lifecycle emissions. Η οριζόντια διακεκομμένη γραμμή υποδεικνύει την πιο πρόσφατη διαθέσιμη τιμή (315,1 g CO₂/kWh, 2025).

Επεκταμένη οπτική (1993-2025). Το Σχήμα 4.18 επεκτείνει τη χρονοσειρά μέχρι το 2025, τοποθετώντας την περίοδο μελέτης σε ευρύτερο ιστορικό πλαίσιο και αναδεικνύοντας την επιτάχυνση της απολιγνιτοποίησης μετά το 2012. Ενώ κατά την περίοδο 1993–2012 ο συντελεστής μειώθηκε κατά ~20% (877 → 704 g CO₂/kWh), στην επόμενη δεκαετιρία (2013–2025) σημείωσε πρόσθετη μείωση ~55% (704 → 315 g CO₂/kWh), φτάνοντας μια συνολική μείωση 64,1% στο διάστημα 1993–2025. Το εύρημα αποτυπώνει τη μεταβατική περίοδο που μεσολάβησε με την είσοδο του φυσικού αερίου, την ανάπτυξη των ΑΠΕ και τη σταδιακή απόσυρση του λιγνίτη.



Σχήμα 4.18: Επεκταμένη διαχρονική εξέλιξη του συντελεστή εκπομπών CO₂ για το ελληνικό δίκτυο, 1993-2025. Η σκιασμένη περιοχή υποδεικνύει την περίοδο μελέτης του σεναρίου. Η διακεκομμένη πορτοκαλί γραμμή παρουσιάζει την περίοδο μετά τη μελέτη για context και ανάδειξη της επιταχυνόμενης απολιγνιτοποίησης.

Συγκριτική αποτίμηση των δύο σεναρίων υπολογισμού. Το Σχήμα 4.19 παρουσιάζει τη σύγκριση μεταξύ ιστορικής εκτίμησης (χρονομεταβλητός συντελεστής) και εκτίμησης με τη σύγχρονη τιμή αναφοράς (2025). Η διαφορά μεταξύ των δύο εκτιμήσεων είναι δραματική και αυξάνεται όσο πηγαίνουμε πιο πίσω στο χρόνο: ενώ το 2012 η αναλογία είναι ~2,23×, για το 1993 ανέρχεται σε ~2,78×. Η σκιασμένη περιοχή του γραφήματος αντιπροσωπεύει το μέγεθος του σφάλματος που θα είχε εισαχθεί εάν χρησιμοποιείτο σταθερός σύγχρονος συντελεστής για τον υπολογισμό ιστορικών εκπομπών.



Σχήμα 4.19: σύγκριση μεταξύ ιστορικής εκτίμησης (χρονομεταβλητός συντελεστής) και εκτίμησης με τη σύγχρονη τιμή αναφοράς (2025).

Σύνοψη Αναλυτικών Ευρημάτων. Η ολοκληρωμένη εικόνα της εικοσαετίας 1993-2012 αναδεικνύει τρείς διαρθρωτικές διαπιστώσεις:

- Διπλασιασμός του ενεργειακού αποτυπώματος των αστικών χρήσεων: Το άθροισμα Οικιακής και Εμπορικής κατανάλωσης σχεδόν διπλασιάστηκε, αντικατοπτρίζοντας τη μετάβαση προς μια πιο εντατική κατανάλωση ενέργειας στον αστικό ιστό.
- Σχετική αποδέσμευση της βιομηχανίας από το ενεργειακό αποτύπωμα της πόλης: Η Βιομηχανική κατηγορία υποχώρησε στην τρίτη θέση έως το τέλος της περιόδου.
- Μη γραμμική μείωση του συντελεστή εκπομπών: Η μείωση 20% κατά την περίοδο μελέτης αντικατοπτρίζει μερικές μόνο διαρθρωτικές αλλαγές στο ενεργειακό μείγμα, με την κύρια απολιγνιτοποίηση να συμβαίνει μετά το 2012.

4.3.7 Συμπεράσματα του Σεναρίου

Το τρίτο σενάριο συμπληρώνει την τριάδα μελετών αναδεικνύοντας μια διαφορετική, αλλά εξίσου σημαντική, διάσταση της προτεινόμενης μεθοδολογίας: τη μοντελοποίηση *minimal, open-data workflows* που, παρά την απλότητά τους, παράγουν δείκτες υψηλής πολιτικής χρησιμότητας. Σε αντιδιαστολή με το Σενάριο 2, όπου η πολυπλοκότητα της διαπραγματεύσεως και της αυτόματης επανασχεδίασης αποτελούσε το επίκεντρο, εδώ η αξία έγκειται στη σημασιολογική ιχνηλασιμότητα ενός υπολογισμού που εξαρτάται από ιστορικά μεταβαλλόμενο επιστημονικό παράμετρο, κρίσιμο για την ορθότητα του αποτελέσματος.

Από μεθοδολογική σκοπιά, το σενάριο επιδεικνύει τέσσερα στοιχεία ιδιαίτερης αξίας. Πρώτον, υποστηρίζει τη συμβατότητα του πλαισίου με διαχρονικές αναλύσεις, επεκτείνοντας την ισχύ των DPV/ODRL/MDAT πέρα από τις συγχρονικές μελέτες. Δεύτερον, αναδεικνύει την ανάγκη ενσωμάτωσης χρονικά μεταβαλλόμενων εξωτερικών παραμέτρων στη ροή υπολογισμού, μια προσέγγιση που μπορεί να γενικευθεί σε άλλες περιπτώσεις όπου ο υπολογισμός εξαρτάται από ιστορικά εξελισσόμενες σταθερές. Τρίτον, αναδεικνύει τη μεθοδολογική σημασία της συνέπειας πηγών: η χρήση τόσο της ιστορικής σειράς όσο και της τιμής αναφοράς από την ίδια πηγή (Ember) με την ίδια μεθοδολογία (lifecycle emissions) εξασφαλίζει ότι οι διαχρονικές συγκρίσεις αντικατοπτρίζουν πραγματικές μεταβολές και όχι διαφορές μεθοδολογίας. Τέταρτον, παρέχει ένα μοντέλο για μελλοντική επεκτασιμότητα: η ενσωμάτωση νέων ετών στη

χρονοσειρά συντελεστών είναι τεχνικά τετριμμένη, αλλά σημασιολογικά πλήρως ιχνηλάσιμη μέσω της JSON-LD αποτύπωσης.

Πέραν της μεθοδολογικής συνεισφοράς, τα ποσοτικά ευρήματα του σεναρίου έχουν αυτοτελή πολιτική αξία. Η ανάδειξη της Οικιακής και Εμπορικής χρήσης ως πρωταγωνιστών του ενεργειακού αποτυπώματος της Θεσσαλονίκης συνιστά τεκμηριωμένη βάση για στόχευση πολιτικών ενεργειακής εξοικονόμησης. Η διαρθρωτική μεταβολή που καταγράφεται περί το 2006 αποτελεί σημείο αναφοράς για μελλοντικές μελέτες ενεργειακής μετάβασης. Ιδιαίτερα η ανάδειξη της επιτάχυνσης της απολιγνιτοποίησης μετά το 2012 (από μείωση 20% σε 55% επιπλέον σε δεκατριετία) προσφέρει εμπειρική επιβεβαίωση της αποτελεσματικότητας της ευρωπαϊκής περιβαλλοντικής πολιτικής.

Τέλος, αξίζει να σημειωθούν δύο περιορισμοί που ορίζουν φυσικές προεκτάσεις της μελέτης. Αφενός, τα δεδομένα κατανάλωσης σταματούν στο 2012, ενώ η χρονοσειρά συντελεστών εκτείνεται μέχρι το 2025. Η επικαιροποίηση των στοιχείων κατανάλωσης μέχρι σήμερα θα επέτρεπε την πλήρη αξιολόγηση της επίδρασης της ενεργειακής μετάβασης. Αφετέρου, οι συντελεστές αφορούν εθνικό επίπεδο, ενώ τοπικές διαφοροποιήσεις στο ενεργειακό μείγμα της Βόρειας Ελλάδας (ιδιαίτερα λόγω της εγγύτητας με σταθμούς λιγνίτη στην Πτολεμαΐδα) δεν αποτυπώνονται. Παρά τους περιορισμούς αυτούς, το σενάριο αποτελεί υπόδειγμα ελαφριάς αλλά σημασιολογικά πλούσιας ροής εργασίας και ολοκληρώνει την παρουσίαση της προστιθέμενης αξίας του προτεινόμενου πλαισίου σε διαφορετικούς τύπους αναλυτικών διεργασιών.

Κεφάλαιο 5ο: Αποτελέσματα και Συζήτηση

5.1 Συγκριτική Σύνθεση των Τριών Σεναρίων

Τα αναλυτικά αποτελέσματα κάθε σεναρίου έχουν παρουσιαστεί διεξοδικά στις αντίστοιχες ενότητες του Κεφαλαίου 4 (4.1.6.2, 4.2.8.2, 4.3.6.2). Στην παρούσα ενότητα η εστίαση μετατοπίζεται από την ανακεφαλαίωση των επιμέρους ευρημάτων στη συγκριτική θεώρηση των τριών σεναρίων ως ενιαίας απόδειξης της εφαρμοσιμότητας του πλαισίου DPW. Η τριάδα επιλέχθηκε ώστε να καλύπτει διακριτές τυπολογίες αναλυτικών εργασιών συγχρονική ανάλυση ανοικτών δεδομένων (Σενάριο 1), υβριδικό περιβάλλον ανοικτών και περιορισμένων δεδομένων με δυναμική επανασχεδίαση workflow (Σενάριο 2), και διαχρονική ανάλυση με ενσωμάτωση χρονομεταβαλλόμενων εξωτερικών παραμέτρων (Σενάριο 3) επιτρέποντας τη συστηματική δοκιμή της προτεινόμενης αρχιτεκτονικής σε διαφορετικές συνθήκες πολυπλοκότητας.

Από συγκριτική σκοπιά αναδεικνύονται τρεις κοινοί άξονες: (α) σημασιολογική ομοιογένεια: παρά τις δομικές διαφορές, και τα τρία σενάρια υπακούουν στο ίδιο σύστημα αντιστοίχισης DPV/ODRL/MDAT, αποδεικνύοντας ότι το λεξιλόγιο είναι επαρκώς εκφραστικό για διαφορετικούς τύπους δεδομένων· (β) προσαρμοζόμενη πολυπλοκότητα: το πλαίσιο εισάγει επιπλέον κόμβους και ελέγχους μόνο εκεί όπου το απαιτούν οι πολιτικές χρήσης (αυτόματη επανασχεδίαση στο Σενάριο 2, απουσία πρόσθετων κόμβων στα Σενάρια 1 και 3)· (γ) πολιτική χρησιμότητα των δεικτών: σε όλες τις περιπτώσεις παράγονται αποτελέσματα άμεσα αξιοποιήσιμα από φορείς αστικής διακυβέρνησης (περιοχές χαμηλού πρασίνου, ζώνες αυξημένης ρυπαντικής έκθεσης, διαρθρωτικές μεταβολές ενεργειακού αποτυπώματος).

Ο Πίνακας 6.1 συνοψίζει συγκριτικά τα κύρια χαρακτηριστικά των τριών σεναρίων, αποτυπώνοντας τη διαφοροποίηση τους ως προς πολυπλοκότητα, τύπους δεδομένων, ενεργοποίηση μηχανισμών Compliance-by-Design και τύπο παραγόμενης γνώσης.

Διάσταση	Σενάριο 1 (Αστικό Πράσινο)	Σενάριο 2 (Ποιότητα Αέρα)	Σενάριο 3 (Ενέργεια & CO ₂)
Τυπολογία ανάλυσης	Συγχρονική, ομαλή ροή	Συγχρονική, υβριδική ροή	Διαχρονική (1993–2012)
Καθεστώς πρόσβασης δεδομένων	Ανοικτά (ODbL)	Μικτό (ODbL + Negotiated)	Ανοικτά (ODbL + Ember)
Αρχικά βήματα workflow	4	5	3
Αυτόματα εισαχθέντες κόμβοι συμμόρφωσης	0	2 (delete + store)	0
Ενεργοποίηση Negotiation Plugin	Όχι	Ναι	Όχι
Παράλληλη εκτέλεση	Ναι (βήματα 1–2)	Ναι (βήματα 1–2)	Όχι (σειριακό)
Τύπος παραγόμενου δείκτη	Στατικός δείκτης κάλυψης	Σύνθετος δείκτης έκθεσης	Διαχρονικός δείκτης CO ₂

Πίνακας 5.1: Συγκριτική θεώρηση των τριών σεναρίων ως προς τα κύρια χαρακτηριστικά εκτέλεσης και τον τύπο των παραγόμενων δεικτών.

5.2 Compliance-by-Design: Θεωρητική Εφαρμογή

Η αρχιτεκτονική φιλοσοφία του Compliance-by-Design έχει αναλυθεί διεξοδικά στην ενότητα 3.2. Στην παρούσα ενότητα η εστίαση μετατοπίζεται από τη θεωρητική θεμελίωση στα εμπειρικά μετρήσιμα ευρήματα που προέκυψαν από την εφαρμογή της στα τρία σενάρια.

Η αρχή παρουσιάστηκε ευκρινέστερα στο Σενάριο 2, όπου η μετάβαση από αμιγώς ανοικτά σε υβριδικό περιβάλλον δεδομένων ενεργοποίησε τους μηχανισμούς αυτόματης επανασχεδίασης. Συγκεκριμένα, ο μηχανισμός επαλήθευσης πολιτικής του consumer Dashboard ανίχνευσε στο ODRL policy του dataset D3 (mdat:ExtendedPopulationData) δύο διακριτές υποχρεώσεις: μη-αποθήκευση των ενδιάμεσων ευαίσθητων δεδομένων (odrl:Prohibition με odrl:action odrl:store) και υποχρεωτική διαγραφή κατόπιν χρήσης (odrl:Obligation με dpv:Erase). Σε απόκριση εισήγαγε αυτόματα δύο πρόσθετους κόμβους στη ροή: έναν dpv:Erase αμέσως μετά την παραγωγή των συγκεντρωτικών δεικτών (βήμα 4*) και έναν dpv:Store στο τέλος της ροής για τη διατήρηση μόνο των τελικών αποτελεσμάτων (βήμα 7*). Συνολικά, η αρχική 5-βηματική ροή μετατράπηκε σε 7-βηματική χωρίς καμία παρέμβαση του Researcher ή του Data Analyst.

Η αντιπαραβολή με τα Σενάρια 1 και 3 είναι εξίσου ενδεικτική. Στα δύο αυτά σενάρια, όπου εμπλέκονται αποκλειστικά ανοικτά δεδομένα χωρίς ρητούς περιορισμούς χρήσης πέραν της απόδοσης πηγής, ο ίδιος μηχανισμός επαλήθευσης δεν εισήγαγε κανέναν πρόσθετο κόμβο. Το γεγονός αυτό υποδηλώνει ότι η πρόσθετη πολυπλοκότητα ενεργοποιείται μόνο εκεί όπου υπάρχουν αντίστοιχες δηλωμένες υποχρεώσεις στα ODRL policies, ικανοποιώντας έτσι την αρχή της αναλογικής επιβολής περιορισμών.

Από την οπτική της διαδρομής ελέγχου, και στα τρία σενάρια κάθε εκτέλεση παρήγαγε πλήρες σημασιολογικό αποτύπωμα σε μορφή JSON-LD, καταγράφοντας τις διεργασίες επεξεργασίας (dpv:Processing) που εκτελέστηκαν, τις ενεργοποιημένες πολιτικές, τους τυχόν αυτόματα εισαχθέντες κόμβους και τα παραγόμενα παραδοτέα. Αυτή η δομημένη τεκμηρίωση καθιστά εφικτό τον εκ των υστέρων έλεγχο συμμόρφωσης από εποπτικές αρχές χωρίς κατά περίπτωση ανασκόπηση κώδικα απαίτηση που εναρμονίζεται με το Άρθρο 30 του GDPR περί καταγραφής δραστηριοτήτων επεξεργασίας.

5.3 Επαναχρησιμοποίηση και Επεκτασιμότητα

Ένα από τα θεμελιώδη πλεονεκτήματα του πλαισίου DPW είναι η ρητή διάκριση μεταξύ της λογικής στρώσης (η τοπολογία των βημάτων, οι σημασιολογικές αντιστοιχίσεις, οι ODRL πολιτικές) και της στρώσης πόρων (τα συγκεκριμένα datasets). Η αρχή αυτή δημιουργεί τις τεχνικές προϋποθέσεις για επαναχρησιμοποίηση, αν και πρέπει να σημειωθεί ότι στα τρία σενάρια η επαναχρησιμοποίηση τεκμηριώθηκε ως δυνητική δυνατότητα και όχι ως εμπειρικά μετρημένο αποτέλεσμα δεν εκτελέστηκε π.χ. στιγμιότυπο επαναφοράς του ίδιου workflow με νέα δεδομένα υπό συνθήκες παραγωγής. Οι σχεδιαστικές δυνατότητες επαναχρησιμοποίησης εκφράζονται διαφορετικά σε κάθε σενάριο. Το πρώτο σενάριο σχεδιάστηκε ώστε να υποστηρίζει την αντικατάσταση των datasets πληθυσμού ή δέντρων με αντίστοιχα νεότερων ετών (π.χ. Απογραφή 2031) ή ομόλογα δεδομένα άλλων δήμων, χωρίς δομικές αλλαγές στη λογική ροή. Το δεύτερο εισάγει την επιπρόσθετη απαίτηση επαναδιαπραγμάτευσης των ODRL policies σε περίπτωση αλλαγής στο dataset περιορισμένης πρόσβασης, ωστόσο η ίδια η ροή και οι σημασιολογικές αντιστοιχίσεις παραμένουν αμετάβλητες. Το τρίτο επιτρέπει τη συστηματική επικαιροποίηση με νεότερες χρονοσειρές κατανάλωσης ή εναλλακτικούς συντελεστές εκπομπών, με την προϋπόθεση ότι διατηρείται η σημασιολογική τους ταυτοποίηση ως mdat:EnergyDataset και mdat:GridCarbonIntensitySeries αντίστοιχα. Η ενσωμάτωση των μεταδεδομένων στα φύλλα Metadata των παραδοτέων Excel και στα συνοδευτικά JSON-LD αρχεία προβλέπεται να προσφέρει αναπαραγωγιμότητα. Η εμπειρική αποτίμηση αυτής της δυνατότητας π.χ. μέσω συστηματικής σύγκρισης αναπαραγωγής ίδιου workflow από τρίτη ομάδα αποτελεί φυσική συνέχεια της παρούσας εργασίας.

5.4 Συμπληρωματικότητα των Σεναρίων

Πέρα από τη συγκριτική αποτύπωση του Πίνακα 5.1, αξίζει να επισημανθεί ότι τα τρία σενάρια δεν επιλέχθηκαν τυχαία, αλλά συγκροτούν μια συμπληρωματική τριάδα. Η σχέση τους δεν είναι αθροιστική αλλά κλιμακούμενη: το Σενάριο 1 αποτελεί το «αρχιτεκτονικό όριο της απλότητας», ένα μια απλή αλλά ολοκληρωμένη ροή εργασίας που υποστηρίζει ότι το πλαίσιο λειτουργεί χωρίς επιπλέον πολυπλοκότητα ακόμη και στις πιο τετριμμένες περιπτώσεις. Το Σενάριο 2 ωθεί την αρχιτεκτονική στο όριο της σύνθετης διακυβέρνησης, αναδεικνύοντας τους μηχανισμούς αυτόματης επανασχεδίασης. Το Σενάριο 3 εισάγει μια επιπλέον μεταβλητή, τον χρόνο δοκιμάζοντας τη συμπεριφορά του πλαισίου παρουσία εξωτερικών παραμέτρων που μεταβάλλονται διαχρονικά.

Η αλληλοκάλυψη των τριών σεναρίων είναι εξίσου σημαντική: όλα μοιράζονται την ίδια γεωγραφική εστίαση (Δήμος Θεσσαλονίκης), αξιοποιούν το ίδιο οικοσύστημα δεδομένων (TDS) και υπόκεινται στην ίδια μεθοδολογική πειθαρχία του MDAT Pilot. Αυτή η κοινή βάση επιτρέπει την ασφαλή σύγκριση των αποτελεσμάτων χωρίς διασταυρούμενες επιδράσεις, ενώ η διαφοροποίηση στη φύση των δεδομένων και των ροών εξασφαλίζει ότι η αποτίμηση του πλαισίου δεν περιορίζεται σε μία στενή κατηγορία αναλυτικών εργασιών.

5.5 Περιορισμοί

Παρά τα θετικά αποτελέσματα, αναγνωρίζονται τέσσερις βασικοί περιορισμοί που οριοθετούν την παρούσα εργασία και αναδεικνύουν τις φυσικές κατευθύνσεις μελλοντικής έρευνας. Χειρωνακτική σημασιολογική ευθυγράμμιση. Η αντιστοίχιση νέων datasets στις κλάσεις και ιδιότητες της οντολογίας MDAT εξακολουθεί να απαιτεί παρέμβαση από εξειδικευμένο χρήστη. Παρότι το πλαίσιο εξασφαλίζει αυτοματοποιημένη συμμόρφωση κατά την εκτέλεση, η αρχική σύνδεση των δεδομένων με την οντολογία παραμένει σημείο τριβής, ιδιαίτερα όταν

τα νέα datasets προέρχονται από ετερογενείς πηγές με ασυνεπή ονοματολογία ή ελλιπή μεταδεδομένα. Γεωγραφικός και χρονικός περιορισμός των σεναρίων. Τα τρία σενάρια αφορούν αποκλειστικά τη Θεσσαλονίκη, γεγονός που περιορίζει τη γενίκευση των ευρημάτων σε διαπεριφερειακό επίπεδο. Επιπλέον, οι χρονικές καλύψεις των δεδομένων δεν είναι πάντοτε ευθυγραμμισμένες (π.χ. στο Σενάριο 3 τα δεδομένα κατανάλωσης σταματούν στο 2012, ενώ η χρονοσειρά συντελεστών εκτείνεται έως το 2025), περιορισμός που οφείλεται κυρίως στη διαθεσιμότητα των ανοικτών δεδομένων. Απουσία συστηματικής αξιολόγησης από τελικούς χρήστες. Η παρούσα εργασία επικυρώνει το πλαίσιο σε επίπεδο τεχνικής λειτουργικότητας και μεθοδολογικής συνέπειας, αλλά δεν περιλαμβάνει εμπειρική αξιολόγηση χρηστικότητας από τελικούς χρήστες (στελέχη δήμων, αναλυτές πολιτικής, υπεύθυνους περιβαλλοντικού σχεδιασμού). Η ποσοτικοποίηση οφελών όπως ο χρόνος σχεδίασης ενός workflow ή ο βαθμός κατανόησης των παραγόμενων δεικτών από μη-τεχνικούς χρήστες παραμένει ανοιχτό ερευνητικό ζήτημα. Έλλειψη σύγκρισης με baseline προσέγγιση. Δεν εκτελέστηκε ποσοτική σύγκριση μεταξύ της προτεινόμενης workflow-driven προσέγγισης και ισοδύναμων συμβατικών scripts ως προς μετρήσιμα κριτήρια (χρόνος εκτέλεσης, ποιότητα παραγόμενων δεικτών, ακρίβεια συμμόρφωσης). Μια τέτοια σύγκριση θα παρείχε ποσοτική επιβεβαίωση της προστιθέμενης αξίας του πλαισίου. Η μελλοντική εργασία θα εστιάσει σε τρεις βασικές κατευθύνσεις: πρώτον, στην ανάπτυξη μηχανισμών αυτόματης σημασιολογικής αντιστοίχισης βασισμένων σε AI, ώστε νέα datasets να συνδέονται με την MDAT οντολογία με ελάχιστη ανθρώπινη παρέμβαση· δεύτερον, στην επέκταση του πλαισίου σε διαπεριφερειακά και διασυνοριακά σενάρια, όπου η ποικιλομορφία των νομικών απαιτήσεων και των δομών δεδομένων θα δοκιμάσει σε μεγαλύτερο βαθμό τα όρια της προτεινόμενης αρχιτεκτονικής· και τρίτον, στη διενέργεια εμπειρικής μελέτης χρηστικότητας με συμμετοχή τελικών χρηστών από φορείς αστικής διακυβέρνησης.

Κεφάλαιο 6ο: Συμπεράσματα και Μελλοντική Εργασία

6.1 Σύνοψη της Συνεισφοράς

Η παρούσα εργασία παρουσίασε και εφάρμοσε πιλοτικά ένα πλαίσιο Data Processing Workflow (DPW), όπως αυτό υλοποιείται στο οικοσύστημα UPCAST, με στόχο τον μετασχηματισμό κατακερματισμένων ανοικτών αστικών δεδομένων σε δομημένους, πολιτικά συμμορφούμενους και αναπαραγωγίμους δείκτες. Η ενσωμάτωση των προτύπων RDF, ODRL 2.2 και DPV 2.2, σε συνδυασμό με την οντολογία MDAT που αναπτύχθηκε ειδικά για τις ανάγκες των σεναρίων μελέτης, σε εκτελέσιμες pipelines κατέδειξε ότι οι σημασιολογικές τεχνολογίες μπορούν να συμβάλουν ουσιαστικά στη γεφύρωση του χάσματος μεταξύ τεχνικής ανάλυσης και νομικής διακυβέρνησης, όχι ως θεωρητική κατασκευή, αλλά ως λειτουργικός μηχανισμός που διαμορφώνει ενεργά τη ροή εκτέλεσης σε ελεγχόμενο πιλοτικό περιβάλλον. Όπως αναδείχθηκε μέσω της πιλοτικής αξιολόγησης στο Κεφάλαιο 6, η προσέγγιση αυτή επιτρέπει την ταυτόχρονη επίτευξη τεσσάρων συμπληρωματικά συνδεδεμένων χαρακτηριστικών αναπαραγωγιμότητα, αυτόματη συμμόρφωση, επαναχρησιμοποίηση και παραγωγή τεκμηριωμένων δεικτών τα οποία συνήθως εμφανίζονται ως αλληλοσυγκρουόμενοι στόχοι σε παραδοσιακές αρχιτεκτονικές ανάλυσης. Πέρα από τη μεθοδολογική διάσταση, η εφαρμογή στα τρία σενάρια της Θεσσαλονίκης παρήγαγε δείκτες με αξία για την τοπική αυτοδιοίκηση, από τη χαρτογράφηση του αστικού πρασίνου και τις ζώνες αυξημένης ρυπαντικής έκθεσης έως τη διαχρονική εξέλιξη του ενεργειακού αποτυπώματος. Η έρευνα οργανώθηκε γύρω από τρία κεντρικά ερευνητικά ερωτήματα που τέθηκαν αρχικά στην ενότητα 1.3, με το πρώτο εξ αυτών (*RQ1*) να εξετάζει πώς μπορεί η σημασιολογική περιγραφή δεδομένων να μετατραπεί από παθητικό μεταδεδωμένο σε ενεργό μηχανισμό διαμόρφωσης ροών εργασίας. Η εργασία υποστηρίζει ότι η μετατόπιση αυτή επιτυγχάνεται μέσω της συστηματικής αντιστοίχισης κάθε βήματος επεξεργασίας με κλάσεις DPV και actions ODRL, υποστηριζόμενη από το εξειδικευμένο λεξιλόγιο πεδίου της οντολογίας MDAT. Η σημασιολογική αυτή υποδομή αποτελεί δομικό στοιχείο το οποίο το Consumer Dashboard ερμηνεύσει για να καθορίσει τη ροή εκτέλεσης· η πιλοτική αξιολόγηση στο Σενάριο 2 ανέδειξε ότι, όταν οι ODRL πολιτικές είναι σημασιολογικά εκφρασμένες, το σύστημα δύναται να τις αναγνώσει αυτόνομα και να ενεργοποιήσει τις αντίστοιχες ενέργειες συμμόρφωσης χωρίς ανθρώπινη παρέμβαση. Αναφορικά με το δεύτερο ερώτημα (*RQ2*), σχετικά με το σε ποιο βαθμό μπορεί η αρχή του Compliance-by-Design να αυτοματοποιήσει τη νομική και ηθική συμμόρφωση μέσω της δυναμικής επανασχεδίασης ροών εργασίας, το σύστημα εισήγαγε αυτόνομα δύο κόμβους συμμόρφωσης (dpv:Erase και dpv:Store) στη ροή του Σεναρίου 2, μετατρέποντας την αρχική 5-βηματική σχεδίαση σε 7-βηματική ροή χωρίς καμία παρέμβαση του Researcher ή του Data Analyst. Η ένδειξη αυτή υποδηλώνει ότι η αυτοματοποίηση συγκεκριμένων υποχρεώσεων συμμόρφωσης είναι εφικτή όταν οι πολιτικές είναι κωδικοποιημένες σε μηχανικά αναγνώσιμη μορφή, αν και ο βαθμός αυτοματοποίησης οριοθετείται τόσο από το γεγονός ότι η αρχική σημασιολογική ευθυγράμμιση των datasets με την οντολογία αναφοράς παραμένει χειρωνακτική διαδικασία, όσο και από το ότι οι μηχανισμοί επανασχεδίασης δοκιμάστηκαν σε προ-ορισμένους τύπους υποχρεώσεων, με τη γενίκευσή τους σε ευρύτερα νομικά πλαίσια να απαιτεί περαιτέρω έρευνα. Τέλος, ως προς το τρίτο ερώτημα (*RQ3*) για τον βαθμό στον οποίο το προτεινόμενο πλαίσιο είναι επεκτάσιμο και επαναχρησιμοποιήσιμο σε διαφορετικούς τύπους αναλυτικών σεναρίων, η εφαρμογή σε τρία σενάρια διαφορετικής τυπολογίας αμιγώς ανοικτά δεδομένα στο Σενάριο 1, υβριδικό

περιβάλλον ανοικτών και περιορισμένων δεδομένων στο Σενάριο 2, και διαχρονική ανάλυση με χρονομεταβαλλόμενες εξωτερικές παραμέτρους στο Σενάριο 3 υποδεικνύει ότι το πλαίσιο μπορεί να ανταποκριθεί σε ετερογενείς αναλυτικές απαιτήσεις χρησιμοποιώντας το ίδιο σύστημα σημασιολογικής αντιστοίχισης. Η επαναχρησιμοποίηση τεκμηριώθηκε ως σχεδιαστική δυνατότητα μέσω της ρητής διάκρισης μεταξύ λογικής στρώσης και στρώσης πόρων, ωστόσο δεν αξιολογήθηκε εμπειρικά με την εφαρμογή υφιστάμενου workflow σε νέα παραγωγικά datasets, καθιστώντας την ποσοτική επικύρωση της επαναχρησιμοποίησης μια φυσική και αναγκαία προέκταση της παρούσας εργασίας.

6.2 Θεωρητικές Συνέπειες

Πέραν της λειτουργικής αξιολόγησης που παρουσιάστηκε στο Κεφάλαιο 6, τα ευρήματα της εργασίας υποδεικνύουν τρεις ευρύτερες θεωρητικές συνέπειες για το πεδίο των δεδομενοκεντρικών αρχιτεκτονικών. Πρώτον, η διάκριση μεταξύ λογικής στρώσης και στρώσης πόρων αναδεικνύεται σε γενικεύσιμη αρχή σχεδιασμού και όχι σε μια απλή τεχνική επιλογή υλοποίησης. Η εφαρμογή σε τρία σενάρια διαφορετικής τυπολογίας υποδηλώνει ότι η αρχή αυτή μπορεί να μεταφερθεί σε άλλους τομείς πέραν της αστικής διακυβέρνησης (όπως η υγεία ή η ενεργειακή έρευνα), όπου η ροή ανάλυσης παραμένει σχετικά σταθερή ενώ οι πόροι δεδομένων ανανεώνονται συνεχώς. Δεύτερον, η συμπληρωματικότητα μεταξύ γενικών (DPV, ODRL) και ειδικευμένων (MDAT) σημασιολογικών πόρων αναδεικνύεται ως πρακτική ανάγκη. Τα γενικά πρότυπα παρέχουν το αναγκαίο υπόβαθρο για τη νομική και ηθική διακυβέρνηση, αλλά η αναλυτική σαφήνεια του εκάστοτε πεδίου απαιτεί ένα εξειδικευμένο λεξιλόγιο. Η συνύπαρξη των δύο επιπέδων μέσω διακριτών namespaces αποτελεί, κατά την κρίση της εργασίας, την αρχιτεκτονικά ενδεδειγμένη λύση. Τρίτον, η αρχή Compliance-by-Design, όπως αναδείχθηκε μέσω της πιλοτικής εφαρμογής στο Σενάριο 2, επεκτείνεται φυσικά σε ευρύτερες απαιτήσεις διακυβέρνησης. Ο ίδιος μηχανισμός που εντοπίζει την κλάση `dpv:RestrictedData` και εισάγει κόμβους `odrl:delete` δύναται να επεκταθεί ώστε να καλύπτει νέους τύπους συμμόρφωσης (όπως η διαχείριση συγκατάθεσης, ο περιορισμός του σκοπού επεξεργασίας και η ελαχιστοποίηση δεδομένων) χωρίς να απαιτείται τροποποίηση της θεμελιώδους αρχιτεκτονικής του συστήματος.

6.3 Τοποθέτηση στο Ευρωπαϊκό Πλαίσιο

Σε ένα ευρωπαϊκό περιβάλλον όπου ο Data Governance Act, ο Data Act και η αρχιτεκτονική των κλαδικών Data Spaces μεταβάλλουν ριζικά τους όρους κοινής αξιοποίησης δεδομένων, η ανάγκη για πλαίσια που ενσωματώνουν τη συμμόρφωση εκ κατασκευής γίνεται όλο και πιο πιεστική. Η παρούσα εργασία τοποθετείται σε αυτό ακριβώς το σημείο, προτείνοντας ένα λειτουργικό μοντέλο εφαρμόσιμο σε ευρωπαϊκή κλίμακα Data Space ecosystems. Η επιλογή θεμελίωσης σε ώριμα ευρωπαϊκά πρότυπα (DPV από το W3C Community Group, ODRL από το W3C) δεν είναι τυχαία: εξασφαλίζει τη σημασιολογική διαλειτουργικότητα του πλαισίου με υφιστάμενες και αναδυόμενες υποδομές, διευκολύνοντας τη μελλοντική μεταφορά του από το ερευνητικό πλαίσιο σε παραγωγικά περιβάλλοντα μεγάλης κλίμακας.

6.4 Μελλοντικές Κατευθύνσεις

Οι περιορισμοί που εντοπίστηκαν στο Κεφάλαιο 5 ορίζουν τρεις φυσικές κατευθύνσεις μελλοντικής έρευνας, καθεμία από τις οποίες αντιμετωπίζει διαφορετικό τύπο πρόκλησης. AI-driven σημασιολογική αντιστοίχιση [23]. Η ανάπτυξη μηχανισμών που αξιοποιούν Large

Language Models ή ειδικά εκπαιδευμένα μοντέλα για την αυτόματη αντιστοίχιση των datasets υπάρχουσες οντολογίες ή σε custom οντολογίες όπως για παράδειγμα η MDAT αντιμετωπίζει το βασικό σημείο τριβής στην εισαγωγή νέων δεδομένων. Η πρόκληση εδώ δεν είναι μόνο τεχνική (ακρίβεια) αλλά και επιστημολογική: πώς διασφαλίζεται ότι η αυτόματη αντιστοίχιση διατηρεί την ίδια σημασιολογική αυστηρότητα με τη χειρωνακτική. Επέκταση σε διαπόλεις και διατομεακά σενάρια. Η μεταφορά του πλαισίου σε σενάρια που εμπλέκουν πολλαπλούς δήμους ή τομείς πολιτικής (π.χ. συνδυασμός υγείας και περιβάλλοντος σε πολυεπίπεδη διακυβέρνηση) θα δοκιμάσει τη γενίκευση των αρχών σε σύνθετα νομικά και οργανωτικά περιβάλλοντα. Η κατεύθυνση αυτή είναι ιδιαίτερα κρίσιμη για την εφαρμογή σε ευρωπαϊκή κλίμακα Data Spaces. Ενσωμάτωση real-time data streams. Η μετάβαση από batch αναλύσεις ιστορικών δεδομένων σε ζωντανές ροές δεδομένων (π.χ. αισθητήρες περιβαλλοντικών παραμέτρων, smart metering) θα επιτρέψει την κατασκευή dashboards αστικής διακυβέρνησης ικανών να υποστηρίξουν λήψη αποφάσεων σε πραγματικό χρόνο. Η πρόκληση εδώ αφορά τη συμβατότητα της σημασιολογικής αρχιτεκτονικής με streaming τεχνολογίες και τη διαχείριση εξελισσόμενων πολιτικών συμμόρφωσης σε δυναμικά περιβάλλοντα.

6.5 Επίλογος

Η επιτυχής εφαρμογή του πλαισίου σε τρία σενάρια αστικής πολιτικής υποδεικνύει ότι η δεδομένο-κεντρική διακυβέρνηση είναι εφικτή, όχι ως οραματικός στόχος, αλλά ως άμεσα υλοποιήσιμη πρακτική. Σε ένα ψηφιακό οικοσύστημα όπου ο όγκος, η ποικιλομορφία και η νομική περιπλοκότητα των δεδομένων αυξάνονται συνεχώς, πλαίσια που ενσωματώνουν τη σημασιολογία ως ενεργό αρχιτεκτονικό στοιχείο και όχι ως διακοσμητικό μεταδεδομένο αποτελούν, κατά την κρίση της εργασίας, την πιο βιώσιμη διαδρομή προς υπεύθυνη και αποτελεσματική αξιοποίηση δημόσιων δεδομένων.

Βιβλιογραφία

- [1] “Data growth worldwide 2029,” Statista. Accessed: May 19, 2026. [Online]. Available: <https://www.statista.com/statistics/871513/worldwide-data-created/>
- [2] “Data Centers - Database.” Accessed: May 21, 2026. [Online]. Available: <https://www.datacentermap.com/datacenters/>
- [3] S. Gheisari *et al.*, “Core Services for Compliant and Trustworthy Data Marketplaces: the EU UPCASt Project”.
- [4] Kovach A, Montalvillo L, Lanza J, Sotres P, and Urbietta A, “Understanding data spaces: A systematic mapping study of foundations technical building blocks and sectoral adoption,” *Comput. Sci. Rev.*, 2026, doi: 10.1016/j.cosrev.2025.100819.
- [5] W. Slabbinck, J. Rojas Meléndez, B. Esteves, P. Colpaert, and R. Verborgh, “Interoperable Interpretation and Evaluation of ODRL Policies,” in *The Semantic Web*, vol. 15719, E. Curry, M. Acosta, M. Poveda-Villalón, M. Van Erp, A. Ojo, K. Hose, C. Shimizu, and P. Lisena, Eds., in Lecture Notes in Computer Science, vol. 15719. , Cham: Springer Nature Switzerland, 2025, p. 19. doi: 10.1007/978-3-031-94578-6_11.
- [6] “REGULATION (EU) 2016/ 679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL - of 27 April 2016 - on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/ 46/ EC (General Data Protection Regulation)”.
- [7] “Data Privacy Vocabularies and Controls Community Group.” Accessed: May 19, 2026. [Online]. Available: <https://www.w3.org/community/dpvcg/>
- [8] P. Moosmann, R. A. Deshmukh, C. Lange, and J. Theissen-Lipp, “Representing Knowledge in Dataspaces”.
- [9] “ODRL Information Model 2.2.” Accessed: May 19, 2026. [Online]. Available: <https://www.w3.org/TR/odrl-model/>
- [10] Μ. Βουράκη, *MariannaVouraki/Mdat-Ontology-definement*. (Apr. 30, 2026). Accessed: May 26, 2026. [Online]. Available: <https://github.com/MariannaVouraki/Mdat-Ontology-definement>
- [11] H. Pettenpohl, M. Spiekermann, and J. R. Both, “International Data Spaces in a Nutshell,” in *Designing Data Spaces*, B. Otto, M. Ten Hompel, and S. Wrobel, Eds., Cham: Springer International Publishing, 2022, pp. 29–40. doi: 10.1007/978-3-030-93975-5_3.
- [12] A. Kovach, L. Montalvillo, J. Lanza, P. Sotres, and A. Urbietta, “Understanding data spaces: A Systematic Mapping Study of foundations, technical building blocks, and sectoral adoption,” *Comput. Sci. Rev.*, vol. 59, p. 100819, Feb. 2026, doi: 10.1016/j.cosrev.2025.100819.
- [13] “Αρχική | Thessaloniki Data Space.” Accessed: May 19, 2026. [Online]. Available: <https://tds.okfn.gr/>
- [14] C. Bizer, T. Heath, and T. Berners-Lee, “Linked Data - The Story So Far,” in *Linking the World’s Information*, 1st ed., O. Seneviratne and J. Hendler, Eds., New York, NY, USA: ACM, 2023, pp. 115–143. doi: 10.1145/3591366.3591378.
- [15] C. Bratsas, E. K. Anastasiadis, A. K. Angelidis, L. Ioannidis, R. Kotsakis, and S. Ougiaroglou, “Knowledge Graphs and Semantic Web Tools in Cyber Threat Intelligence: A Systematic Literature Review,” *J. Cybersecurity Priv.*, vol. 4, no. 3, pp. 518–545, Sep. 2024, doi: 10.3390/jcp4030025.

- [16] A. Cimmino, J. Cano-Benito, and R. García-Castro, “Open Digital Rights Enforcement framework (ODRE): From descriptive to enforceable policies,” *Comput. Secur.*, vol. 150, p. 104282, Mar. 2025, doi: 10.1016/j.cose.2024.104282.
- [17] E. I. Papagiannakopoulou, N. L. Dellas, G. V. Lioudakis, M. N. Koukovini, and A. Mousas, “Semantic conflict resolution for access and usage control”.
- [18] P. Fafalios, Y. Marketakis, A. Axaridou, Y. Tzitzikas, and M. Doerr, “A Workflow Model for Holistic Data Management and Semantic Interoperability in Quantitative Archival Research,” *Digit. Scholarsh. Humanit.*, vol. 38, no. 3, pp. 1049–1066, Aug. 2023, doi: 10.1093/llc/fqad018.
- [19] G. Marchment, S. Cohen-Boulakia, and F. Lemoine, “Computational Reproducibility With Scientific Workflows: Analysing viral genomes with Nextflow,” in *ACM REP’25 ACM Conference on Reproducibility and Replicability*, Vancouver, Canada, Apr. 2025. Accessed: May 19, 2026. [Online]. Available: <https://hal.science/hal-05525260>
- [20] J. Chandra and S. K. Navneet, “Policy-Driven AI in Dataspaces: Taxonomy, Explainability, and Pathways for Compliant Innovation,” Jul. 30, 2025, *arXiv*: arXiv:2507.20014. doi: 10.48550/arXiv.2507.20014.
- [21] B. F. Sveinbjarnarson, “Scaling Data Governance Through Automation and Dynamic Data Spaces”.
- [22] Efremidis O, Verlekis D, Koutras I, Lioudakis G, Papagiannakopoulou E, and Efremidis O, “D4.1 – UPCAST Integrated Architecture,” 2025. [Online]. Available: <https://www.upcast-project.eu/wp-content/uploads/2025/12/D4.1-UPCAST-Integrated-.pdf>
- [23] Chandra J and Navneet S K, “Policy-Driven AI in Dataspaces,” *CoRR*, 2025, [Online]. Available: <https://arxiv.org/abs/2507.20014>
- [24] “Data Center Map.”
- [25] “GAIA-X Architecture Document,” 2022. [Online]. Available: <https://docs.gaia-x.eu/policy-rules-committee/architecture-document/22.10/>

Παραρτήματα Κεφαλαίων

Τα παρόντα παραρτήματα παρέχουν εκτεταμένες επεξηγήσεις, τεχνικές λεπτομέρειες και συμπληρωματικό υλικό για κάθε κεφάλαιο της εργασίας.

Παράρτημα Α: Επέκταση Κεφαλαίου 1 - Το Παγκόσμιο Τοπίο Παραγωγής Δεδομένων

Α.1 Στατιστικά Παραγωγής Δεδομένων

Για να αντιληφθούμε το μέγεθος της επανάστασης των δεδομένων, είναι χρήσιμο να εξετάσουμε μερικά ενδεικτικά στοιχεία:

Μέτρο	Τρέχουσα Κατάσταση	Πρόβλεψη
Παγκόσμιος όγκος δεδομένων	181 ZB (2025)	394 ZB (2028)
Συνδεδεμένες IoT συσκευές	~18 δισ. (2025)	40,6 δισ. (2034)
Επενδύσεις ψηφιακής μετασχηματισμού	~2 τρισ. \$ (2023)	3,88 τρισ. \$ (2028)
Ενεργά Data Centers παγκοσμίως	11.028	Σε ανάπτυξη
Data Centers στην Ελλάδα	25	—

Πίνακας Α.1: Βασικά στατιστικά στοιχεία παραγωγής δεδομένων

Α.2 Γεωγραφική Κατανομή Data Centers

Η γεωγραφική κατανομή των data centers αντικατοπτρίζει την ψηφιακή ωριμότητα κάθε χώρας. Η ΗΠΑ ηγείται με 4.036 εγκαταστάσεις, ακολουθούμενη από το Ηνωμένο Βασίλειο (511), τη Γερμανία (507) και τη Γαλλία (322). Η Ελλάδα, με 25 εγκαταστάσεις (εκ των οποίων πολλές στη Θεσσαλονίκη), βρίσκεται σε τροχιά ανάπτυξης, υποστηρισμένη από ευρωπαϊκές πρωτοβουλίες όπως το GAIA-X και το Data Act [24], [25].

Α.3 Ελληνικές Πύλες Ανοικτών Δεδομένων

Στην Ελλάδα, η πρόσβαση σε ανοικτά δεδομένα οργανώνεται μέσω:

- data.gov.gr: Εθνικός κατάλογος δεδομένων (10.000+ datasets, 340+ φορείς)
- opendata.ellak.gr: Πύλη ΕΕΛΛΑΚ για ανοικτά πρότυπα
- Thessaloniki Data Space (tds.okfn.gr): 120+ datasets για τη Θεσσαλονίκη
- opendata.businessportal.gr (ΓΕΜΗ): Επιχειρηματικά δεδομένα μέσω API
- IMEdD Lab: Datasets για ερευνητική δημοσιογραφία

Παράρτημα Β: Επέκταση Κεφαλαίου 2 - Τεχνικές Λεπτομέρειες Σημασιολογικών Τεχνολογιών

B.1 RDF Turtle Παράδειγμα για MDAT Dataset

Παρακάτω παρουσιάζεται ένα ενδεικτικό παράδειγμα της αναπαράστασης ενός dataset σε RDF Turtle, χρησιμοποιώντας τα namespaces mdat:, dpv: και dcat: (Data Catalog Vocabulary):

```
< @prefix mdat: <https://github.com/MariannaVouraki/Mdat-Ontology-definement#>
@prefix dpv: <https://w3id.org/dpv#> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
```

```
mdat:UrbanTreeCategoriesDataset a dcat:Dataset, mdat:TreesDistribution ;
    dcat:title "Urban Tree Categories Thessaloniki"@el ;
    dcat:publisher <https://tds.okfn.gr/> ;
    dpv:hasDataCategory dpv:EnvironmentalData ;
    odrl:hasPolicy mdat:ODbL1Policy .
```

```
mdat:ODbL1Policy a odrl:Policy ;
    odrl:permission [ a odrl:Permission ;
        odrl:action odrl:use, odrl:derive, odrl:display ] ;
    odrl:obligation [ a odrl:Obligation ;
        odrl:action odrl:attribute ] .
```

Αυτή η αναπαράσταση επιτρέπει στο consumer Dashboard να ανακτήσει αυτόματα τις επιτρεπόμενες ενέργειες και να επαληθεύσει ότι κάθε βήμα του workflow καλύπτεται.

B.2 ODRL Policy Παράδειγμα για Restricted Dataset

Για το extended demographic dataset του Σεναρίου 2, η διαπραγματευθείσα πολιτική εκφράζεται σε ODRL ως εξής:

```
mdat:RestrictedDemographicPolicy a odrl:Policy ;
    odrl:permission [ a odrl:Permission ;
        odrl:action odrl:use ;
        odrl:constraint [ odrl:leftOperand odrl:purpose ;
            odrl:operator odrl:eq ;
            odrl:rightOperand dpv:NonCommercialResearch ] ] ;
    odrl:prohibition [ a odrl:Prohibition ;
```

odrl:action odrl:store] ;
odrl:obligation [a odrl:Obligation ;
odrl:action odrl:delete ;
odrl:target mdat:ExtendedPopulationData] .

Το Consumer Dashboard διαβάζει αυτή την πολιτική και εισάγει αυτόματα τον κόμβο *odrl:delete* μετά το βήμα 3 του workflow.

B.3 Σύγκριση Λειτουργιών

Λειτουργία	Περιγραφή	Χρήση στα Σενάρια
dpv:Collect	Συλλογή δεδομένων από πηγή	Εισαγωγή datasets από TDS
dpv:Transform	Μετασχηματισμός/καθαρισμός	CleanAndNormalise
dpv:Aggregate	Συγκέντρωση/ομαδοποίηση	Υπολογισμός μέσων τιμών ρύπων
dpv:Derive	Παραγωγή νέων δεδομένων	Υπολογισμός δεικτών
dpv:Visualise	Οπτικοποίηση	GenerateGraphsAndVisualSummaries
dpv:Analyse	Ανάλυση	AssessComplianceWithEULimits
odrl:delete	Διαγραφή (υποχρεωτική)	Αυτόματος κόμβος dpv:delete
dpv:Store	Αποθήκευση (ελεγχόμενη)	Αυτόματος κόμβος dpv:store

Πίνακας B.1: DPV Processing Operations - Σύγκριση και Χρήση στα Σενάρια

Παράρτημα Γ: Επέκταση Κεφαλαίου 3 - Τεχνικές Λεπτομέρειες DPW

Γ.1 Αρχιτεκτονική Nextflow

Η μηχανή Nextflow χρησιμοποιείται για την εκτέλεση των DPW ως πλήρως αναπαραγώγιμα computational pipelines. Η Nextflow εκφράζει κάθε βήμα ως ανεξάρτητη "process", με ρητά ορισμένες εισόδους και εξόδους:

```
process CleanAndNormalise {  
    input: path energy_dataset  
    output: path "cleaned_energy.csv"  
    script:  
        """  
        python clean_normalise.py ${energy_dataset} cleaned_energy.csv  
        """  
}
```

Η Nextflow διαχειρίζεται αυτόματα τον παραλληλισμό (π.χ. τα βήματα 1 και 2 του Σεναρίου 1 εκτελούνται ταυτόχρονα), την επανεκκίνηση σε περίπτωση σφάλματος, και την καταγραφή αποτελεσμάτων.

Γ.2 Συνδυασμός ODRL + DPV στη Πράξη: Βήμα-Βήμα

Το πλαίσιο επαλήθευσης πολιτικής λειτουργεί ως εξής πριν από κάθε εκτέλεση workflow:

5. Ανάκτηση ODRL policy κάθε dataset από το TDS SPARQL endpoint
6. Εξαγωγή permission/prohibition/obligation triples
7. Αντιστοίχιση κάθε workflow βήματος με τη dpv:Processing operation
8. Επαλήθευση ότι κάθε dpv:Processing έχει αντίστοιχο odrl:Permission
9. Εντοπισμός odrl:Obligation και αυτόματη εισαγωγή κόμβων compliance
10. Δημιουργία audit log με πλήρη τεκμηρίωση για κάθε απόφαση

Γ.3 Σύγκριση Σεναρίων: Χαρακτηριστικά DPW

Χαρακτηριστικό	Σενάριο 1	Σενάριο 2	Σενάριο 3
Πλήθος datasets	2	2 (1 restricted)	1
Πλήθος κόμβων (αρχικό)	4	5	3
Αυτόματα εισαχθέντα βήματα	0	2 (delete+store)	0
Διαπραγμάτευση	Όχι	Ναι	Όχι
Τύπος δεδομένων	Δημόσια	Δημόσια + Restricted	Δημόσια
Παράλληλη εκτέλεση	Ναι (βήματα 1+2)	Ναι (βήματα 1+2)	Όχι (σειριακό)

Αδειοδότηση	ODbL 1.0	ODbL + Negotiated	ODbL 1.0
-------------	----------	-------------------	----------

Πίνακας Γ.1: Συγκριτική Ανάλυση Χαρακτηριστικών DPW ανά Σενάριο

Παράρτημα Δ: Επέκταση Κεφαλαίου 5 - Αναλυτικά Αποτελέσματα

Δ.1 Αποτελέσματα Σεναρίου 2: Ρύπανση ανά Δημοτική Κοινότητα

Δημοτική Κοινότητα	NO (μg/m ³)	NO ₂ (μg/m ³)	PM ₁₀ (μg/m ³)	PM _{2.5} (μg/m ³)	Δείκτης έκθεσης/κάτοικο	Αξιολόγηση
1η Δημοτική Κοινότητα	108,9 	85,6 	49,4 	34,5 	0,007246	Υψιστη Προτεραιότητα
2η Δημοτική Κοινότητα	48,6	39,1	47,3 	—	—	Υψηλή Προτεραιότητα
3η Δημοτική Κοινότητα	11,6	—	30,1	20,3	—	Κανονικό
4η Δημοτική Κοινότητα	—	—	32,7	—	—	Παρακολούθηση
5η Δημοτική Κοινότητα	—	—	—	21,6	—	Κανονικό
Όριο ΕΕ	100	40	40	25	—	—

Πίνακας Δ.1: Μέσες συγκεντρώσεις ρύπων ανά δημοτική κοινότητα Θεσσαλονίκης (2010–2013).

 = υπέρβαση ορίου ΕΕ

Δ.2 Τύπος Υπολογισμού CO₂ Εκπομπών (Σενάριο 3)

Ο τύπος που εφαρμόζεται στο βήμα **ComputeCO₂Emissions** για κάθε κατηγορία χρήσης *c* και έτος *y*:

$$\text{CO}_2 \text{ (kg)}_{y,c} = \frac{\text{Consumption}_{y,c} \text{ (kWh)} \times \text{EF}_y \text{ (g CO}_2\text{/kWh)}}{1000}$$

όπου EF_y ο συντελεστής έντασης άνθρακα του ελληνικού δικτύου για το έτος *y*. Η διαίρεση με 1000 μετατρέπει το αποτέλεσμα από γραμμάρια σε κιλά CO₂.

Πηγή και Μεθοδολογία Συντελεστών

Οι συντελεστές προέρχονται από την επίσημη βάση δεδομένων της **Ember Climate** (*Yearly Electricity Data, 2025*), διαθέσιμη μέσω του *Our World in Data*. Υπολογίζονται με μεθοδολογία **lifecycle emissions**, η οποία συμπεριλαμβάνει:

- τις άμεσες εκπομπές από την καύση καυσίμων,
- τις upstream εκπομπές μεθανίου από την αλυσίδα εφοδιασμού,
- τις εκπομπές που σχετίζονται με την κατασκευή και λειτουργία των σταθμών παραγωγής.

Όλες οι τιμές εκφράζονται σε **ισοδύναμα CO₂ σε ορίζοντα 100 ετών (CO₂eq, GWP100)**. Η χρήση της ίδιας πηγής και μεθοδολογίας τόσο για την ιστορική σειρά (1993–2012) όσο και για την τιμή αναφοράς (2025) εξασφαλίζει μεθοδολογική συνέπεια στις διαχρονικές συγκρίσεις και αποτρέπει την εισαγωγή τεχνητών διαφορών οφειλόμενων σε ετερογένεια μεθόδων υπολογισμού.

Δ.3 Πηγές Δεδομένων TDS

Dataset	Άδεια	MDAT Κλάση	Χρήση
Urban Tree Categories	ODbL 1.0	mdat:TreesDistribution	Σενάριο 1
Permanent Population Census 2011	ODbL 1.0	mdat:PopulationDistribution	Σενάριο 1 & 2
Air Quality Measurements 2010–2013	ODbL 1.0	mdat:EnvironmentalData	Σενάριο 2
Extended Population Data	Negotiated	mdat:ExtendedPopulationData	Σενάριο 2
Energy Consumption 1993–2012	ODbL 1.0	mdat:EnergyDataset	Σενάριο 3

Πίνακας Δ.2: Datasets που χρησιμοποιήθηκαν ανά σενάριο

Παράρτημα Ε: Γλωσσάριο Τεχνικών Όρων

Όρος	Ορισμός
Compliance-by-Design	Αρχιτεκτονική φιλοσοφία όπου η νομική συμμόρφωση ενσωματώνεται από τον σχεδιασμό, όχι εκ των υστέρων
DAG	Directed Acyclic Graph: Κατευθυνόμενος ακυκλικός γράφος που αναπαριστά τις εξαρτήσεις μεταξύ βημάτων ενός workflow
Data Space	Αποκεντρωμένο, ομοσπονδιακό οικοσύστημα δεδομένων που εξασφαλίζει data sovereignty για τους παρόχους
DPV	Data Privacy Vocabulary: W3C πρότυπο για σημασιολογική έκφραση GDPR-συμβατών επεξεργασιών δεδομένων
DPW	Data Processing Workflow: Δομημένη, εκτελέσιμη ακολουθία βημάτων επεξεργασίας με σημασιολογική τεκμηρίωση
Knowledge Graph	Γράφος γνώσης: Δίκτυο οντοτήτων και σχέσεων σε RDF μορφή
Linked Data	Βέλτιστες πρακτικές για δημοσίευση και σύνδεση δομημένων δεδομένων στον Ιστό μέσω URIs
MDAT	Custom οντολογία (namespace: mdat:) που αναπτύχθηκε για τα σενάρια της παρούσας εργασίας
Nextflow	Μηχανή εκτέλεσης computational workflows με υποστήριξη παραλληλισμού και αναπαραγωγιμότητας
ODRL	Open Digital Rights Language: W3C πρότυπο για μηχανικώς αναγνώσιμες πολιτικές χρήσης δεδομένων
Ontology (Οντολογία)	Επίσημος, τυπικός ορισμός εννοιών και σχέσεων ενός πεδίου γνώσης σε μηχανικώς αναγνώσιμη μορφή
OWL	Web Ontology Language: Γλώσσα έκφρασης οντολογιών με πλούσιους εκφραστικούς μηχανισμούς
Pipeline	Αλληλουχία επεξεργαστικών βημάτων όπου η έξοδος ενός βήματος τροφοδοτεί το επόμενο
RDF	Resource Description Framework: Βασικό μοντέλο δεδομένων του Semantic Web, βασισμένο σε τριάδες (Subject-Predicate-Object)
Semantic Web	Ο "Σημασιολογικός Ιστός": Επέκταση του WWW όπου δεδομένα εκφράζονται με τρόπο κατανοητό τόσο από ανθρώπους όσο και από μηχανές
SPARQL	Γλώσσα ερωτημάτων για RDF δεδομένα, ανάλογη με SQL για σχεσιακές βάσεις
TDS	Thessaloniki Data Space: Πλατφόρμα ανοικτών δεδομένων για τη Θεσσαλονίκη (https://tds.okfn.gr/)
Triple (Τριάδα)	Στοιχειώδης μονάδα πληροφορίας στο RDF: Subject – Predicate – Object
UPCAST	Universal Platform Components for Safe Fair Interoperable Data Exchange, Monetization and Trading
URI	Uniform Resource Identifier: Παγκοσμίως μοναδικό αναγνωριστικό για οντότητες στον Ιστό
Workflow Template	Αποθηκευμένο, επαναχρησιμοποιήσιμο πρότυπο workflow που μπορεί να εφαρμοστεί σε νέα datasets

Παράρτημα ΣΤ: Πλατφόρμες Δεδομένων και Εργαλεία

ΣΤ.1 Ελληνικές Πύλες Ανοικτών Δεδομένων

Πλατφόρμα	URL	Κατηγορία
Data.gov.gr	https://data.gov.gr	Κυβερνητικά δεδομένα
OpenData.ellak.gr	https://opendata.ellak.gr	Ανοικτά πρότυπα
Thessaloniki Data Space	https://tds.okfn.gr	Αστικά δεδομένα
BusinessPortal OpenData	https://opendata.businessportal.gr	Επιχειρηματικά

Πίνακας ΣΤ.1: Ελληνικές Πύλες Ανοικτών Δεδομένων

ΣΤ.2 Εμπορικές Πλατφόρμες Δεδομένων

Πλατφόρμα	Χαρακτηριστικά	Τομέας
Nokia Data Marketplace	Datasets AI/ML για τηλεπικοινωνίες	Telecom, Smart Cities
AWS Data Exchange	Marketplace με δεκάδες παρόχους	Γενικό
Snowflake Marketplace	Cloud-native, χωρίς μεταφορά δεδομένων	Analytics, ML
Dawex Data Exchange	B2B ανταλλαγή datasets	Πολλαπλοί κλάδοι
Google Cloud Marketplace	APIs και datasets για ML	AI, Analytics

Πίνακας ΣΤ.2: Εμπορικές Πλατφόρμες Δεδομένων

ΣΤ.3 Εργαλεία που Χρησιμοποιήθηκαν

Εργαλείο	Ρόλος	Τεχνολογία
ABOVO Consumer Dashboard	Σχεδιασμός και εκτέλεση workflows	Node-based visual editor
Nextflow	Εκτέλεση computational pipeline	Groovy/Bash, Docker
Python	Αναλυτική επεξεργασία δεδομένων	pandas, numpy, matplotlib
MDAT Ontology	Σημασιολογική αντιστοίχιση	OWL/RDF (Turtle)
ODRL/DPV	Πολιτικές χρήσης / Ιδιωτικότητα	W3C Recommendations
UPCAST Discovery Plugin	Αναζήτηση datasets	SPARQL, REST API

UPCAST Negotiation Plugin	Διαπραγμάτευση πρόσβασης	Digital contracts
------------------------------	--------------------------	-------------------

Πίνακας ΣΤ.3: Εργαλεία που Χρησιμοποιήθηκαν στην Παρούσα Εργασία