ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΣΤΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

Διπλωματική Εργασία

ΔΗΜΟΣΙΑ ΑΝΟΙΧΤΑ ΔΕΔΟΜΕΝΑ:

ΠΡΩΤΟΒΟΥΛΙΕΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ

του

ΑΘΑΝΑΣΑΚΗ ΓΕΩΡΓΙΟΥ

ΑΜ: 1210

Επιβλέποντες καθηγητές:

Ταραμπάνης Κωνσταντίνος, Καθηγητής

Ταμπούρης Ευθύμιος, Επίκουρος Καθηγητής

1ος Εξεταστής: Μαργαρίτης Κωνσταντίνος, Καθηγητής

2ος Εξεταστής: Κίτσιος Φώτιος, Λέκτορας

Υποβλήθηκε ως προαπαιτούμενο για την απόκτηση του μεταπτυχιακού

διπλώματος ειδίκευσης στα Πληροφοριακά Συστήματα

Σεπτέμβριος 2012

Copyright © Αθανασάκης Γιώργος, 2011-2012

Αυτή η εργασία χορηγείται με άδεια Creative Commons Αναφορά Δημιουργού 3.0 Ελλάδα. Για χρήση αυτής της εργασίας είναι απαραίτητη η αναφορά στον συγγραφέα. Η συγκεκριμένη άδεια επιτρέπει την ελεύθερη χρήση αυτής της εργασίας για Διανομή (αναπαραγωγή, διανομή, παρουσίαση στο κοινό του έργου) και Διασκευή (για να τροποποιήσετε το έργο). Για περισσότερες πληροφορίες σχετικά με την άδεια της εργασίας επισκεφθείτε τον σύνδεσμο http://creativecommons.org/licenses/by/3.0/gr/ .

Η έγκριση της μεταπτυχιακής εργασίας από το Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών στα Πληροφοριακά Συστήματα του Πανεπιστημίου Μακεδονίας δεν υποδηλώνει απαραιτήτως και αποδοχή των απόψεων του συγγραφέα εκ μέρους των Τμημάτων (Ν.5343/32 αρ.202 παρ.2).

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα αρχικά να ευχαριστήσω αυτούς που μου συμπαραστάθηκαν ώστε να καταφέρω να τελειώσω αυτήν την εργασία δηλαδή την Σίσσυ την Αννούλα και την Σήλια που με ανέχθηκαν και πολλές φορές ακόμη και χωρίς να το ξέρουν βοήθησαν πάρα πολύ, τους γονείς, τον Δημήτρη και την Χαρούλα, τον Αντρέα και την Γεωργία και τέλος όλους τους συναδέλφους στην εταιρεία «Αφοι Χρυσομάλλη ΑΕ» με ιδιαίτερη αναφορά στον Παλαιολόγο.

Επίσης θα ήθελα να ευχαριστήσω τους καθηγητές κ. Ταραμπάνη και κ. Ταμπούρη για τις κατευθύνσεις και τις συμβουλές τους και τέλος τον κ. Καλαμπόκη για την βοήθεια και την υπομονή του να δίνει απαντήσεις και να συζητάει τις περίεργες απορίες μου.

Ιδιαίτερες ευχαριστίες ανήκουν στο σύνολο των συναδέλφων συμφοιτητών για την βοήθειά τους και την συμπαράστασή τους σε όλη την διάρκεια αυτού του ωραίου ταξιδιού.

ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα 1

Κατάλογος Εικόνων 4

Κατάλογος Πινάκων 6

Επιτελική Σύνοψη 7

Executive summary 9

1. Εισαγωγή 11

1.1. Περιγραφή Προβλήματος 11

1.2. Αντικείμενο και Στόχοι Μελέτης 11

1.3. Περιεχόμενα Μελέτης 13

2. Μεθοδολογία 15

2.1. Εισαγωγή 15

2.2. Ανοιχτά Δημόσια Δεδομένα στο Διαδίκτυο 15

2.3. Μεθοδολογία Χαρτογράφησης των Βρετανικών Ανοιχτών Δημόσιων Δεδομένων 17

2.4. Υλοποίηση Ερευνας και Οπτικοποίηση 18

3. Ανασκόπηση βιβλιογραφίας 20

3.1. Εισαγωγή 20

3.2. Ηλεκτρονική και Ανοιχτή Διακυβέρνηση 21

3.2.1. Εισαγωγή 21

3.2.2. Ηλεκτρονική Διακυβέρνηση και Μοντέλα Ανάπτυξης 22

3.2.3. One-stop e-government 23

3.2.4. Από την Ηλεκτρονική Διακυβέρνηση στην Ανοιχτή Διακυβέρνηση 24

3.2.5. Ανοιχτή Διακυβέρνηση 25

3.3. Δημόσια Ανοιχτά Δεδομένα 26

3.3.1. Εισαγωγή 26

3.3.2. Η Κοινωνία της Πληροφορίας 27

3.3.3. Ορισμοί 29

3.3.4. Αρχές των Ανοιχτών Κυβερνητικών Δεδομένων 32
3.3.5. Χρησιμότητα Δημοσιοποίησης των Ανοιχτών Κυβερνητικών Δεδομένων 36
3.4. Επαναχρησιμοποίηση των Ανοιχτών Κυβερνητικών Δεδομένων 38
3.4.1. Εισαγωγή 38
3.4.2. Ευρωπαϊκή Ενωση και Νομικό Πλαίσιο Επαναχρησιμοποίησης 38
3.4.3. Εμπορική Αξία των Ανοιχτών Δημόσιων Δεδομένων 41
3.4.4. Μορφές Επαναχρησιμοποιούμενων Δεδομένων 44
3.5. Τρόποι Δημοσίευσης των Ανοιχτών Κυβερνητικών Δεδομένων 47
3.5.1. Εισαγωγή 47
3.5.2. Πρακτικές Δημοσιοποίησης 47
3.5.3. Συνδεδεμένα (Κυβερνητικά) Ανοιχτά Δεδομένα 48
3.5.4. Κατάλογοι Δημοσίων Κυβερνητικών Δεδομένων 52
3.5.5. Εφαρμογές που Χρησιμοποίούν Δημοσια Κυβερνητικά Δεδομένα 59
3.6. Εμπόδια και Προβλήματα στην Υλοποίηση των Δημόσιων Κυβερνητικών Δεδομένων 62
3.6.1. Εισαγωγή 62
3.6.2. Προκλήσεις 63
3.6.3. Η Ποιότητα της Πληροφορίας και η Χρήση της ως Εμπόρευμα 64
4. Επισκόπηση της Βρετανικής Ιστοσελίδας Ανοιχτών Δημόσιων Δεδομένων 67
4.1. Εισαγωγή 67
4.2. Ιστορικά Στοιχεία 67
4.3. Δομή και Περιεχόμενο 70
5. Αποτελέσματα ερευνας Συνόλων Δεδομένων 72
5.1. Εισαγωγή 72
5.2. Διαδικασία της Ερευνας 72
5.2.1. Σχεδιασμός και Προετοιμασία 72
5.2.2. Συλλογή Δεδομένων 73
5.2.3. Αξιολόγηση Δεδομένων 75

5.2.4. Κατηγοριοποίηση Δεδομένων 80

5.2.5. Καθορισμός των Σχέσεων των Συνόλων Δεδομένων 84

5.2.6. Ανάλυση Αποτελεσμάτων Ερευνας 96

5.2.7. Οπτικοποίηση Δεδομένων 105

6. Συμπεράσματα και Προτάσεις για Μελλοντική Ερευνα 107

Βιβλιογραφία 112

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 2.1 Μεθοδολογία έρευνας 18

Εικόνα 3.1 Δεδομένα με τη συμβατική άποψη σχετικά με την ιεραρχία της γνώσης 30

Εικόνα 3.2 Το οικοσύστημα των ανοιχτών δεδομένα 32

Εικόνα 3.3 Χρονοδιάγραμμα επιλεγμένων γεγονότων σχετικά με τα Δημόσια Κυβερνητικά Δεδομένα 40

Εικόνα 3.4 Διάγραμμα συνδεδεμένων συνόλων δεδομένων τον Μάρτιο 2009 51

Εικόνα 3.5 Διάγραμμα συνδεδεμένων συνόλων δεδομένων τον Σεπτέμβριο 2011..... 51

Εικόνα 3.6 Παγκόσμιος χάρτης πρωτοβουλιών Δημόσιων Κυβερνητικών Δεδομένων 55

Εικόνα 3.7 Οπτικοποίηση δημοσίων δαπανών Ηνωμένου Βασιλείου ανά γενική κατηγορία .60

Εικόνα 3.8 Οπτικοποίηση δημοσίων δαπανών Ηνωμένου Βασιλείου ανά περιοχή 60

Εικόνα 3.9 Διεπαφή αποστολής προβλήματος μαζί με απεικόνιση σε χάρτη 61

Εικόνα 4.1 Beta έκδοση της ιστοσελίδας 68

Εικόνα 4.2 Ιστοσελίδα http://www.data.gov.uk πριν τον Ιούνιο του 2012 69

Εικόνα 4.3 Ιστοσελίδα http://www.data.gov.uk μετά τον Ιούνιο του 2012 69

Εικόνα 4.4 Εργαλείο οπτικοποίησης οργανογραμμάτων των υπουργείων 70

Εικόνα 4.5 Μορφές αρχείων 71

Εικόνα 5.1 Εισαγωγή αρχείου μεταδεδομένων στο Google Refine 74

Εικόνα 5.2 Παράδειγμα συνόλου δεδομένων με πολλά αρχεία..... 76

Εικόνα 5.3 Παράδειγμα συνόλου δεδομένων με ένα αρχείο..... 78

Εικόνα 5.4 Γράφημα γενικών κατηγοριών 80

Εικόνα 5.5 Γράφημα γεωγραφικών οντοτήτων..... 82

Εικόνα 5.6 Ομαδοποίηση συνόλων δεδομένων..... 83

Εικόνα 5.7 Ιεραρχική απεικόνιση κατηγοριοποίησης των δεδομένων με εκκίνηση την γενική κατηγορία 83

Εικόνα 5.8 Ιεραρχική απεικόνιση κατηγοριοποίησης των δεδομένων με εκκίνηση την γεωγραφική οντότητα 84

Εικόνα 5.9 Οπτικοποίηση πρώτου βήματος προσέγγισης με εκκίνηση την γενική κατηγορία 85

Εικόνα 5.10 Οπτικοποίηση δεύτερου βήματος προσέγγισης με εκκίνηση την γενική κατηγορία 86

Εικόνα 5.11 Οπτικοποίηση τρίτου βήματος προσέγγισης με εκκίνηση την γενική κατηγορία. 86

Εικόνα 5.12 Οπτικοποίηση πρώτου βήματος προσέγγισης με εκκίνηση την γεωγραφική οντότητα 87

Εικόνα 5.13 Οπτικοποίηση δεύτερου βήματος προσέγγισης με εκκίνηση την γεωγραφική οντότητα 87

Εικόνα 5.14 Οπτικοποίηση τρίτου βήματος προσέγγισης με εκκίνηση την γεωγραφική οντότητα 88

Εικόνα 5.15 Οπτικοποίηση Γεωγραφικών οντοτήτων 97

Εικόνα 5.16 Οπτικοποίηση γενικών κατηγοριών 100

Εικόνα 5.17 Οπτικοποίηση κατηγορίας «Κυβέρνηση, Πολιτική και Δημόσια Διοίκηση» 102

Εικόνα 5.18 Οπτικοποίηση οντοτήτων 104

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 2.1 Σχετικές περιπτώσεις για διαφορετικές ερευνητικές μεθόδους 18

Πίνακας 3.1 Κατηγορίες Ανοιχτών Δημόσιων Δεδομένων με παραδείγματα 43

Πίνακας 3.2 Πιθανές μορφές δεδομένων για Ανοιχτά Κυβερνητικά Δεδομένα 46

Πίνακας 3.3 Πρότυπα διάθεσης δεδομένων 46

Πίνακας 3.4 Σχήμα κατάταξης πρωτοβουλιών Δ.Κ.Δ. 54

Πίνακας 3.5 Κατάταξη υπαρχόντων πρωτοβουλιών σύμφωνα με το σχήμα 54

Πίνακας 3.6 Πίνακας πρωτοβουλιών Δ.Κ.Δ σε παγκόσμιο επίπεδο 58

Πίνακας 5.1 Κομμάτι τελικού πίνακα συνόλων δεδομένω 79

Πίνακας 5.2 Κομμάτι του πίνακα Annual Survey of hours and earnings 91

Πίνακας 5.3 Κομμάτι του πίνακα Personal Incomes from tax records 92

Πίνακας 5.4 Κομμάτι του πίνακα Biological quality of rivers and canals 93

Πίνακας 5.5 Κομμάτι του πίνακα Chemical quality of rivers and canals 94

Πίνακας 5.6 Κομμάτι των πινάκων των δύο datasets, πυκνότητα εργασίας και αριθμός καταγεγραμμένων διαρρήξεων. 95

Πίνακας 5.7 Αριθμός συνόλων δεδομένων και γενική ταξινόμηση 96

Πίνακας 5.8 Αριθμός συνόλων δεδομένων και ταξινόμηση ανά κατηγορία 98

Πίνακας 5.9 Ανάλυση υποκατηγοριών δεδομένων κατηγορίας «Κυβέρνηση, Πολιτική και Δημόσια Διοίκηση» 101

Πίνακας 5.10 Ανάλυση οντοτήτων συνόλων δεδομένων 103

ΕΠΙΤΕΛΙΚΗ ΣΥΝΟΨΗ

Ο Δημόσιος Τομέας είναι υπεύθυνος για την παραγωγή ενός μεγάλου όγκου δεδομένων τα οποία αφορούν τις κυβερνητικές διαδικασίες και το κοινωνικό σύνολο. Η ανοιχτή διακυβέρνηση ως νέο μοντέλο διακυβέρνησης, στηρίζεται στην δημοσιοποίηση των δεδομένων που συγκεντρώνει ο δημόσιος τομέας, ώστε αυτά να είναι ανοιχτά στους πολίτες και στις επιχειρήσεις και γενικά στον οποιονδήποτε θα ήθελε να τα μελετήσει και να τα χρησιμοποιήσει. Η επαναχρησιμοποίηση των δεδομένων από επαγγελματίες ή ιδιώτες αλλά και από τους ίδιους τους κυβερνητικούς οργανισμούς μπορεί να προσθέσει αξία στα δεδομένα που κάθονται ανεκμετάλλευτα στους σκληρούς δίσκους των υπολογιστών των δημόσιων οργανισμών.

Ο στόχος της παρούσας έρευνας είναι να μελετηθούν τα σύνολα δεδομένων του Βρετανικού καταλόγου που διατηρεί τα ανοιχτά δεδομένα του δημόσιου τομέα και να γίνει μία χαρτογράφηση αυτών, ώστε ο δυνητικός χρήστης να έχει έναν οδηγό εύρεσης και σύνθεσης των δεδομένων. Επισκεφθήκαμε τα 7595 ενεργά σύνολα δεδομένων που εμφανίζονταν στην ιστοσελίδα στις 31 Οκτωβρίου 2011 με σκοπό να τα κατηγοριοποιήσουμε και να βρούμε τις σχέσεις τους μέσα από τις κατηγορίες αυτές. Μετά την αφαίρεση των γεωχωρικών δεδομένων και αυτών που δεν είναι προσβάσιμα, δημιουργήσαμε έναν πίνακα με 4638 στατιστικά σύνολα δεδομένων τα οποία κατατάξαμε σύμφωνα με την γενική τους κατηγορία, την υποκατηγορία τους, την γεωγραφική τους οντότητα και την οντότητα την οποία περιγράφουν. Τα περισσότερα σύνολα δεδομένων και συγκεκριμένα το 36% αυτών που κατηγοριοποιήσαμε αφορούν την δημόσια διοίκηση και πιο συγκεκριμένα τα τμήματα της κυβέρνησης και τους δημόσιους οργανισμούς και το 73% αυτών αφορά τις δαπάνες. Επίσης από τα αποτελέσματα διαπιστώνουμε ότι το 41% των δεδομένων γεωγραφικά περιγράφει την Αγγλία και το 25% το Ηνωμένο Βασίλειο.

Παράλληλα με την έρευνα μελετήσαμε τα δημόσια ανοιχτά δεδομένα με σκοπό να τα ορίσουμε και να παρουσιάσουμε τις αρχές τους, να εξετάσουμε το νομικό πλαίσιο για την επαναχρησιμοποίηση τους και τους τρόπους με τους οποίους αυτά δημοσιοποιούνται. Η τάση προς το «άνοιγμα» των δημοσίων δεδομένων είναι σχετικά καινούργια με αποτέλεσμα να παρουσιάζονται

προβλήματα στην υλοποίηση του από τις κυβερνήσεις. Η μελέτη των προβλημάτων αυτών μαζί με τις λύσεις που προτείνονται είναι κομμάτι της εργασίας. Η βιβλιογραφική επισκόπηση ενίσχυσε την επιλογή της βρετανικής κυβερνητικής ιστοσελίδα παροχής ανοιχτών δεδομένων αναδεικνύοντας την ως την πιο εξελιγμένη και ενημερωμένη.

Μετά την μελέτη των δεδομένων, την κατηγοριοποίηση τους και την ανάλυση των αποτελεσμάτων δημιουργήσαμε δύο οπτικοποιημένα σχήματα με την βοήθεια του εργαλείου οπτικοποίησης bubble tree με σκοπό την διευκόλυνση του χρήστη στην αναζήτηση των δεδομένων. Τα σχήματα οδηγούν τον χρήστη βαθιά στα δεδομένα με κριτήριο τις μεταξύ τους σχέσεις και καταλήγουν σε ομάδες δεδομένων με κοινά χαρακτηριστικά. Ένα σύνολο δεδομένων από μόνο του, πιθανόν να παρέχει αποσπασματικές πληροφορίες ή πληροφορίες που με τον συνδυασμό τους με κάποιες άλλες να δίνουν χρήσιμα εργαλεία για την λήψη αποφάσεων και τα αποτελέσματα της έρευνας μαζί με την οπτικοποίηση προσφέρουν τον χάρτη για την κατάλληλη σύνθεσή τους.

EXECUTIVE SUMMARY

Public Sector is responsible for producing a large volume of data relating to government operations and the community. Open government, as a new governance model, is based on the disclosure of data collected by the public sector so that they are open to citizens and businesses and generally to anyone who wants to study them and use them. Reuse of data by professionals or individuals, but also by government organizations, can add value to the data that remain unused on the hard disks of computers in public agencies.

The goal of this dissertation is to study the datasets that are listed in the British catalogue of open government data and map them in order for the potential user to have a guide for finding and aggregating the data. We visited the 7595 active datasets that were listed in the website on the 31st of October 2011, in order to categorize them and find the relations within these categories. After removing geospatial and inaccessible datasets, we created a table of 4638 statistical datasets and we classified them by a general category, a subcategory, their geographical entity and the entity described. Most datasets, namely 36% of those, were categorized relating to public administration and more specifically government departments and public institutions. The 73% of those datasets are about expenditure. Also the results shown that 41% of data have geographically coverage England and 25% the UK.

During research, we studied open government data in order to define and present their principles, examine the legal framework for the reuse and the ways in which they are made public. The trend towards the "opening" of public data is relatively new hence there are problems in the implementation by governments. The study of these problems, along with the solutions proposed, is part of the dissertation. The literature review supported the choice of the British government open data website and revealed that it is the most sophisticated and up to date of all relative websites.

After data study, categorization and result analysis, we created two schemes visualized with the aid of the visualization tool “bubble tree” in order to facilitate any potential user in searching the data. Schemes lead users deep into data based on the

relations among them and the result is a group of datasets with common characteristics. Datasets themselves are likely to provide incomplete information or information that, combined with others, provides useful tools for decision making and research results along with visualization provide the map for the right aggregation.

1. ΕΙΣΑΓΩΓΗ

1.1. ΠΕΡΙΓΡΑΦΗ ΠΡΟΒΛΗΜΑΤΟΣ

Ο Δημόσιος Τομέας, τα τελευταία χρόνια, δημοσιοποιεί όλο και περισσότερα σύνολα δεδομένων στο διαδίκτυο. Αυτή η τάση ενισχύει την διαφάνεια και την υπευθυνότητα από την πλευρά των κυβερνήσεων και την καινοτομία από την πλευρά των πολιτών και των επιχειρήσεων που θέλουν να εμπλακούν στην επαναχρησιμοποίηση των δεδομένων αυτών. Ενώ οι κυβερνητικοί κατάλογοι δεδομένων βελτιώνονται συνέχεια, στην πραγματικότητα ο χρήστης που ψάχνει σύνολα δεδομένων για επαναχρησιμοποίηση βρίσκεται μπροστά σε έναν μεγάλο αριθμό δεδομένων χωρίς να ξέρει πώς πιθανώς να μπορούσε να συνθέσει πληροφορίες μεταξύ τους.

Τα σύνολα των δεδομένων περιγράφουν συγκεκριμένες πληροφορίες που από μόνες τους υπάρχει πιθανότητα να είναι αποσπασματικές. Ο συνδυασμός τους όμως με άλλα σύνολα δεδομένων, τα οποία περιγράφουν τις ίδιες οντότητες ή έχουν τις ίδιες μεταβλητές, μπορεί να προσφέρει ένα πολύτιμο εργαλείο στα χέρια αυτών που θέλουν να χρησιμοποιήσουν τα δεδομένα. Οι Shadbolt et al. (2010) σημειώνουν χαρακτηριστικά ότι ανακαλύπτοντας τα καταλληλότερα σημεία ένωσης για διαφορετικά σύνολα δεδομένων, τα οποία είναι σημεία αναφοράς που τα δεδομένα έχουν κοινά, αυτά είναι πολύτιμα για την σύνθεση τους.

Στα πλαίσια της συγκεκριμένης έρευνας θα προσπαθήσουμε να σχετίσουμε τα σύνολα δεδομένων που είναι δημοσιευμένα στην βρετανική ιστοσελίδα παροχής ανοιχτών κυβερνητικών δεδομένων με βάση τις οντότητες και τις μεταβλητές που περιγράφουν, προσπαθώντας να βοηθήσουμε τον πιθανό χρήστη να πλοηγηθεί στον ωκεανό των δεδομένων.

1.2. ΑΝΤΙΚΕΙΜΕΝΟ ΚΑΙ ΣΤΟΧΟΙ ΜΕΛΕΤΗΣ

Ο κύριος στόχος της παρούσας εργασίας είναι να μελετηθούν σε βάθος τα ανοιχτά σύνολα δεδομένων της κυβερνητικής ιστοσελίδας της Μεγάλης Βρετανίας (http://data.gov.uk), με σκοπό να τα κατηγοριοποιήσουμε και να διαπιστώσουμε τις μεταξύ τους σχέσεις ώστε να δημιουργηθούν οι προϋποθέσεις με την σύνθεση και

επαναχρησιμοποίηση τους να εξαχθούν χρήσιμα συμπεράσματα. Το κίνημα για το άνοιγμα των κυβερνητικών δεδομένων ξεκίνησε σχετικά πρόσφατα και η δημοσίευση των δεδομένων έγινε αρχικά μαζικά, χωρίς την προηγούμενη επεξεργασία τους για να πληρούν τις αρχές των ανοιχτών δεδομένων. Οι κατάλογοι που δημοσίευσαν τα δεδομένα, δίνουν τρόπους αναζήτησης χωρίς ποιοτικά χαρακτηριστικά και ο χρήστης χάνεται στα δεδομένα χωρίς χάρτη. Ο στόχος μας είναι να δημιουργήσουμε ένα χάρτη για τον χρήστη που αναζητά την σύνθεση των δεδομένων με κοινά χαρακτηριστικά.

Η επιλογή της ιστοσελίδας ενισχύθηκε από την βιβλιογραφική επισκόπηση, η οποία ανέδειξε την συγκεκριμένη ιστοσελίδα ως την πιο συμβατή με τις αρχές των δημόσιων κυβερνητικών δεδομένων. Πιο συγκεκριμένα οι επιμέρους στόχοι της εργασίας είναι οι παρακάτω:

- Η μελέτη των εννοιών της ανοιχτής διακυβέρνησης η οποία βασίζεται στα δημόσια ανοιχτά δεδομένα.
- Η διατύπωση των αρχών των δημόσιων ανοιχτών δεδομένων.
- Η παρουσίαση της έννοιας της επαναχρησιμοποίησης των δεδομένων και πως δημιουργείται προστιθέμενη αξία από αυτή, καθώς και των μορφών των δεδομένων ώστε αυτά να μπορούν να επαναχρησιμοποιηθούν.
- Η εξέταση των τρόπων με τους οποίους τα δεδομένα δημοσιοποιούνται στο διαδίκτυο και γίνονται «ανοιχτά» για τον οποιονδήποτε θέλει να τα χρησιμοποιήσει.
- Η μελέτη των διαπιστωμένων προβλημάτων στην υλοποίηση των ανοιχτών δημόσιων δεδομένων.
- Η μελέτη σε βάθος των συνόλων δεδομένων που υπάρχουν στο http://data.gov.uk .
- Η δημιουργία σχέσεων μεταξύ των οντοτήτων που περιγράφουν τα σύνολα δεδομένων του Βρετανικού portal ώστε να μπορέσουμε μελλοντικά να επαναχρησιμοποιήσουμε τα δεδομένα για την λήψη αποφάσεων.

1.3. ΠΕΡΙΕΧΟΜΕΝΑ ΜΕΛΕΤΗΣ

Στο κεφάλαιο 2 παρουσιάζεται η μεθοδολογία που επιλέχθηκε για την υλοποίηση της εργασίας. Η μεθοδολογία περιλαμβάνει την μελέτη των αρχών των δημόσιων δεδομένων και της υπάρχουσας κατάστασης των πρωτοβουλιών και τεχνολογιών ανά τον κόσμο και την μεθοδολογία για την συλλογή των δεδομένων για την παρουσίαση της πρωτοβουλίας δημοσιοποίησης των κυβερνητικών ανοιχτών δεδομένων του Ηνωμένου Βασιλείου.

Στο κεφάλαιο 3 γίνεται η επισκόπηση της βιβλιογραφίας για τα ανοιχτά δημόσια δεδομένα, ξεκινώντας από την παρουσίαση της ανοιχτής διακυβέρνησης και την σχέση της με τα ανοιχτά δεδομένα. Στην συνέχεια περιγράφουμε τους ορισμούς και τις αρχές σύμφωνα με τις οποίες τα δεδομένα μπορούν να θεωρηθούν ότι είναι ανοιχτά. Η εμπορική αξία που προστίθεται στα δεδομένα από την επαναχρησιμοποίηση τους, το νομικό πλαίσιο της επαναχρησιμοποίησης και οι τρόποι επαναχρησιμοποίησης εξετάζονται παρακάτω στο τρίτο κεφάλαιο και στο τελευταίο μέρος του παρουσιάζονται οι κυβερνητικές πρωτοβουλίες που υπάρχουν παγκόσμια και παραδείγματα εφαρμογών που έχουν δημιουργηθεί και λειτουργούν χρησιμοποιώντας ανοιχτά κυβερνητικά δεδομένα. Το κεφάλαιο κλείνει με τα προβλήματα και τα εμπόδια που έχουν εντοπιστεί αναφορικά με την γραφειοκρατία, την νοοτροπία των δημόσιων οργανισμών και τα τεχνικά κωλύματα.

Στο κεφάλαιο 4 παρουσιάζουμε την ιστοσελίδα http://www.data.gov.uk η οποία είναι η πρωτοβουλία του Ηνωμένου Βασιλείου για τα ανοιχτά δημόσια δεδομένα και λειτουργεί ως κατάλογος για την εύρεσή τους. Ξεκινάει με κάποια ιστορικά στοιχεία και συνεχίζει με την δομή και το περιεχόμενο της ιστοσελίδας. Η εξέταση της γίνεται περιγραφικά καθώς το ενδιαφέρον μας είναι για το περιεχόμενο και όχι για την λειτουργικότητα της.

Στο κεφάλαιο 5 αναλύονται ο σχεδιασμός και η πραγματοποίηση της έρευνας, τα εργαλεία που χρησιμοποιήθηκαν κατά την διάρκεια της, πως έγινε η συλλογή των δεδομένων καθώς και τα αποτελέσματα που τελικά προέκυψαν. Τα αποτελέσματα αναφέρονται στην κατηγοριοποίηση των δεδομένων και πως αυτά σχετίζονται μεταξύ τους. Στο τέλος περιγράφεται ο τρόπος και το εργαλείο που χρησιμοποιήθηκε για την οπτικοποίηση των αποτελεσμάτων της έρευνας.

Στο έκτο και τελευταίο κεφάλαιο καταγράφουμε τα συμπεράσματα από την επισκόπηση της βιβλιογραφίας και την έρευνα και κάνουμε κάποιες προτάσεις για μελλοντική έρευνα.

2. ΜΕΘΟΔΟΛΟΓΙΑ

2.1. ΕΙΣΑΓΩΓΗ

Στο παρόν κεφάλαιο θα παρουσιάσουμε την μεθοδολογία την οποία ακολουθήσαμε στην προσπάθειά μας να μελετήσουμε την παρούσα κατάσταση των Δημόσιων ανοιχτών δεδομένων, των πρωτοβουλιών που υπάρχουν και των τεχνολογιών που χρησιμοποιούνται. Επίσης από την αρχή θέσαμε σαν στόχο την χαρτογράφηση των συνόλων δεδομένων που υπάρχουν στην Βρετανική κυβερνητική ιστοσελίδα http://www.data.gov.uk και θα παρουσιάσουμε την μεθοδολογία προσέγγισης αυτού του εγχειρήματος.

2.2. ΑΝΟΙΧΤΑ ΔΗΜΟΣΙΑ ΔΕΔΟΜΕΝΑ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

Στην προσπάθεια αναζήτησης πηγών για να μπορέσουμε να εκτιμήσουμε την παρούσα κατάσταση των Δημοσίων ανοιχτών δεδομένων την περίοδο που διανύουμε, επικεντρωθήκαμε σε κάποιες ομάδες και συγγραφείς οι οποίοι έχουν μεγάλο ενδιαφέρον το οποίο προκύπτει από την ενασχόλησή τους με το θέμα τα τελευταία χρόνια, καθώς και το θέμα έχει πάρει μεγάλη έκταση την τελευταία πενταετία. Σαν αποτέλεσμα αυτού του γεγονότος η κλασσική βιβλιογραφία μας προσέφερε κάποια πολύ ενδιαφέροντα άρθρα σε επιστημονικά περιοδικά, αλλά βρήκαμε και άλλες ενδιαφέρουσες πηγές οι οποίες ήταν ιστοσελίδες οργανισμών που ασχολούνται με τα ανοιχτά δεδομένα, άρθρα σε ηλεκτρονικές εφημερίδες, βίντεο ομιλητών σε συνέδρια, επίσημες οδηγίες της Ευρωπαϊκής Ένωσης κ.α..

Μια σημαντική πηγή άντλησης πληροφοριών και δεδομένων είναι οι ιστοσελίδες της Ευρωπαϊκής Επιτροπής και της Ευρωπαϊκής Ένωσης καθώς ασχολούνται με τα δημόσια δεδομένα από το 2003 και είναι οι άμεσα υπεύθυνοι για την υλοποίηση του ανοίγματος των δεδομένων στις χώρες που την αποτελούν. Οι αποφάσεις και οι ντιρεκτίβες της Ευρωπαϊκής Επιτροπής βοήθησαν στην κατανόηση της σοβαρότητας με την οποία η Ευρώπη αντιμετωπίζει το θέμα αλλά και της κατάστασης των ανοιχτών δημοσίων δεδομένων στην Ευρώπη σήμερα.

Ο εφευρέτης του world wide web (www) sir Tim Berners-Lee έχει γράψει αρκετές μελέτες για την δημοσίευση των ανοιχτών δημοσίων δεδομένων και για τα συνδεδεμένα δεδομένα. Η αναζήτηση του ονόματός του έφερε πολλά αποτελέσματα συγγραμμάτων τα οποία χρησιμοποιήθηκαν στην παρούσα εργασία και βοήθησαν στην κατανόηση του θέματος καθώς η βαρύτητα του έργου του είναι πολύ μεγάλη. Επίσης η ομάδα ενδιαφέροντος για την ηλεκτρονική διακυβέρνηση World Wide Web Consortium (W3C) της οποίας ηγείται ο Tim Berners-Lee έχει αρκετές αναφορές στα ανοιχτά δημόσια δεδομένα τις οποίες και χρησιμοποιήσαμε για την ολοκλήρωσης της εργασίας.

Το Open Knowledge Foundation (OKF) είναι ένας οργανισμός μη κερδοσκοπικού χαρακτήρα του οποίου τα μέλη ασχολούνται με την ανοιχτή γνώση. Μία ομάδα εργασίας αυτού του οργανισμού έχει ως μοναδική της ενασχόληση την ανοιχτή διακυβέρνηση και τα δημόσια ανοιχτά δεδομένα. Από τα μέλη αυτής της ομάδας εργασίας και τις μελέτες που έχουν δημοσιεύσει μπορέσαμε να αντλήσουμε πηγές για την επισκόπηση και την κατάσταση των δεδομένων αυτή την στιγμή στον κόσμο. Μέλη αυτής της ομάδας εργασίας των οποίων συγγράμματα χρησιμοποιήσαμε στην συγκεκριμένη εργασία είναι ο Alonso Jose, ο Cyganiac Richard, ο Davies Tim, ο Eaves David, ο Dekkers Maxx, οι Ann και Brian Fitzgerald, ο Grey Jonathan, η Janssen Katleen, ο Pollock Rufus και ο Tauberer Josh, τα ονόματα των οποίων εμφανίζονται πολύ ψηλά στις αναζητήσεις για δημόσια ανοιχτά δεδομένα. Επίσης η εγγραφή στην λίστα των e-mail του OKF μας βοήθησε να βρούμε πηγές καθώς και η μελέτη της ανταλλαγής απόψεων μέσω ηλεκτρονικού ταχυδρομείου όλων των παραπάνω μελών του OKF βοήθησε να δούμε πως προσεγγίζουν το θέμα.

Η διερεύνηση στις παραπάνω ομάδες, οργανισμούς και συγγραφείς μας οδήγησε στην εύρεση των κατάλληλων μελετών και συγγραμμάτων έτσι ώστε να περιγράψουμε καλύτερα την κατάσταση των ανοιχτών δημόσιων δεδομένων τον τελευταίο χρόνο παγκόσμια.

2.3. ΜΕΘΟΔΟΛΟΓΙΑ ΧΑΡΤΟΓΡΑΦΗΣΗΣ ΤΩΝ ΒΡΕΤΑΝΙΚΩΝ ΑΝΟΙΧΤΩΝ ΔΗΜΟΣΙΩΝ ΔΕΔΟΜΕΝΩΝ

Ο αρχικός στόχος της έρευνας στην παρούσα εργασία ήταν η χαρτογράφηση των δεδομένων που βρίσκονται διαθέσιμα στο ευρύ κοινό από την ιστοσελίδα http://data.gov.uk και είναι δεδομένα που δημοσιεύουν οι βρετανικοί δημόσιοι οργανισμοί. Η επιλογή των βρετανικών δεδομένων έγινε γιατί η βιβλιογραφία ανέδειξε την βρετανική πρωτοβουλία ως την πιο σύγχρονη και την πιο σύμφωνη με τις τεχνολογίες και τις αρχές των ανοιχτών δημόσιων δεδομένων (Kalampokis et al., 2011a). Παράλληλα με την χαρτογράφηση των δεδομένων θα έπρεπε να συλλέξουμε στοιχεία για την ιστοσελίδα για να μπορέσουμε να καταλάβουμε πως λειτουργεί και τέλος να δείξουμε τις σχέσεις μεταξύ των μεταβλητών και των οντοτήτων που περιγράφουν τα δεδομένα αυτά.

Έχοντας καθορίσει τους στόχους της έρευνας και έχοντας επιλέξει την πηγή που θα συλλέξουμε τα δεδομένα μας προχωρήσαμε με βάση το σχήμα της εικόνας 2.2 στην πραγματοποίηση της έρευνας αντιμετωπίζοντας το θέμα ως μία μελέτη περίπτωσης (case study). Σύμφωνα με τον Yin (2009) η μελέτη περίπτωσης ως μέθοδος προτιμάται όταν:

1. Θέτονται ερωτήματα τύπου «πώς» ή «γιατί»
2. Ο ερευνητής έχει μικρό έλεγχο πάνω στα γεγονότα
3. Εστιάζουμε σε ένα σύγχρονο φαινόμενο μέσα σε ένα πραγματικό πλαίσιο

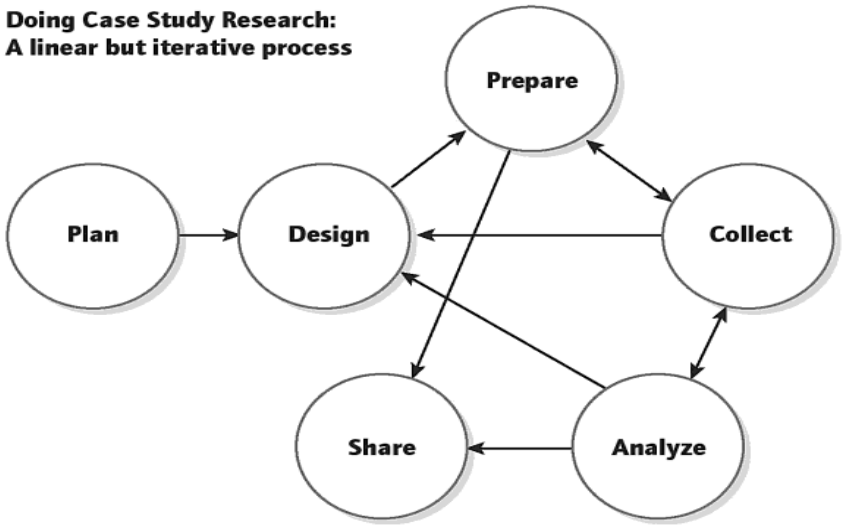
Η επιλογή της μεθοδολογίας της μελέτης περίπτωσης εξαρτάται σε μεγάλο βαθμό από τους στόχους της έρευνας. Όσο περισσότερο οι στόχοι ψάχνουν να εξηγήσουν κάποια παρούσα περίπτωση ή φαινόμενο τόσο περισσότερο η μελέτη περίπτωσης θα είναι σχετικότερη. Αυτές οι τρείς προϋποθέσεις θέτονται στον πίνακα 2.1 για κάθε μέθοδο έρευνας με σκοπό να μας βοηθήσει να επιλέξουμε την κατάλληλη.

Η συλλογή δεδομένων έγινε αποκλειστικά από την ιστοσελίδα με την βοήθεια ενός αρχείου dump το οποίο περιέχει τα μεταδεδομένα των συνόλων δεδομένων του http://www.data.gov.uk σε μία δεδομένη ημερομηνία (31/10/2011), δηλαδή ενός στιγμιότυπου των συνόλων δεδομένων.

METHOD	(1) Form of Research Question	(2) Requires Control of Behavioral Events?	(3) Focuses on Contemporary Events?
Experiment	how, why?	yes	yes
Survey	who, what, where, how many, how much?	no	yes
Archival Analysis	who, what, where, how many, how much?	no	yes/no
History	how, why?	no	no
Case Study	how, why?	no	yes

Figure 1.1 Relevant Situations for Different Research Methods

Πίνακας 2.1 Σχετικές περιπτώσεις για διαφορετικές ερευνητικές μεθόδους (Yin, 2009)



Εικόνα 2.1Μεθοδολογία έρευνας (Yin, 2009)

2.4. ΥΛΟΠΟΙΗΣΗ ΕΡΕΥΝΑΣ ΚΑΙ ΟΠΤΙΚΟΠΟΙΗΣΗ

Ξεκινώντας την διαδικασία της έρευνας ακολουθήσαμε τα βήματα της μεθοδολογίας που επιλέξαμε, δηλαδή της μελέτης περίπτωσης, και αυτά ήταν τα παρακάτω:

1. Σχεδιασμός: Τα δεδομένα υπήρχαν στον κατάλογο (ιστοσελίδα) και στο αρχείο dump υπήρχαν τα μεταδεδομένα τους. Αρχικά θα μελετούσαμε το

αρχείο και στην συνέχεια θα συγκρίναμε τα στοιχεία με τα δεδομένα της ιστοσελίδας και στο τέλος θα τα κατηγοριοποιούσαμε και θα συσχετίζαμε.

2. Προετοιμασία: Επιλέξαμε το εργαλείο google refine για την ανάλυση του αρχείου των μεταδεδομένων και το excel για την αποθήκευση και ανάλυση των δεδομένων.
3. Συλλογή: Συλλέξαμε τα δεδομένα με βάση τον αρχικό σχεδιασμό και αφαιρέσαμε αυτά που δεν χρειαζόμασταν (ανενεργά και γεωχωρικά), χρησιμοποιώντας το google refine και το excel.
4. Ανάλυση: Αξιολογήσαμε τα δεδομένα και ξεκινήσαμε την κατηγοριοποίηση. Μετά την πρώτη ανάλυση κρίναμε αναγκαίο ότι έπρεπε να προσθέσουμε κάποια στοιχεία (υποκατηγοριοποίηση, χρονολογική κάλυψη και εφαρμογή) και επιστρέψαμε στην συλλογή δεδομένων. Τελικά επαναλάβαμε την διαδικασία της ανάλυσης μέχρι να πάρουμε τα τελικά αποτελέσματα.
5. Οπτικοποίηση: Επιλέξαμε το εργαλείο οπτικοποίησης bubble tree visualization το οποίο παραμετροποίησαμε και έπειτα προετοιμάσαμε κατάλληλα τα δεδομένα μας για την σωστή εμφάνιση τους από την εφαρμογή .

3. ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ

3.1. ΕΙΣΑΓΩΓΗ

Η προσπάθεια των κυβερνήσεων να πλησιάσουν τους πολίτες βρήκε αρωγό ένα πολύτιμο εργαλείο, το internet. Η μετατροπή της παραδοσιακής διακυβέρνησης σε ηλεκτρονική ενώ περικλείει διάφορα διαδικαστικά θέματα, όπως πρόσβαση στο διαδίκτυο από την πλειονότητα των πολιτών και τηλεπικοινωνιακές υποδομές από την πλευρά του δημόσιου τομέα, τελικά περνώντας από διάφορα στάδια ανάπτυξης καταλήγει στην παροχή αυτοματοποιημένων υπηρεσιών προς τους πολίτες μέσω του διαδικτύου (Chen, 2009; Layne and Lee, 2001). Οι διαδικτυακές πύλες, one-stop e-government, συγκέντρωσαν σ` έναν δικτυακό τόπο (σημείο, ιστοσελίδα) τις περισσότερες από τις υπηρεσίες που θα μπορούσε το κράτος να προσφέρει στους πολίτες, όμως η εξέλιξη της τεχνολογίας και η ανάγκη για περισσότερη διαφάνεια και συμμετοχή από τους πολίτες στις κυβερνητικές διαδικασίες, οδήγησε σε ένα νέο μοντέλο διακυβέρνησης που ονομάζεται ανοιχτή διακυβέρνηση (Dias and Rafael; 2007; Tapscott, 2010). Το κυριότερο χαρακτηριστικό του νέου αυτού μοντέλου είναι τα Δημόσια Ανοιχτά Δεδομένα.

Μετά από μία σύντομη περιγραφή των μοντέλων της Ηλεκτρονικής διακυβέρνησης στο κεφάλαιο 2.2 και τους λόγους για την ανάγκη μετάβασης στην ανοιχτή διακυβέρνηση, αναφερόμαστε στα Δημόσια Ανοιχτά Δεδομένα στο κεφάλαιο 2.3 εξετάζοντας αρχικά το πως οι νόμοι περί ελευθερίας της πληροφορίας στην Αμερική και στην Ευρώπη οδήγησαν στην ανοιχτή διακυβέρνηση και στο «άνοιγμα» των δεδομένων των κυβερνήσεων (Access to information, 2006; Parycek and Sachs, 2010). Οι απαραίτητοι ορισμοί για το τι σημαίνουν τα Δημόσια Ανοιχτά Δεδομένα ακολουθούν (Open Definition, 2011a) σύμφωνα με το Open Knowledge Foundation, μαζί με τις αρχές τους (8 Principles of Open Government Data, 2007). Το κεφάλαιο κλείνει με την απάντηση στο ερώτημα γιατί πρέπει να τα Δημόσια δεδομένα να είναι ανοιχτά.

Στο κεφάλαιο 2.4 περιγράφεται το πολύ σημαντικό θέμα της επαναχρησιμοποίησης των δεδομένων αρχικά εξετάζοντας την Ευρωπαϊκή οδηγία για τα δεδομένα του δημόσιου τομέα (European Commission 2003) και τις

αναθεωρήσεις της, αλλά και τις μελέτες που έγιναν για την εμπορική αξία που αποκτούν αυτά τα δεδομένα αν επαναχρησιμοποιηθούν (PIRA, 2000; Dekkers et al., 2006; Pollock, 2008, 2011). Στην συνέχεια αφού παραθέτουμε τις κυριότερες κατηγορίες των δεδομένων (Vickery, 2011) βλέπουμε με ποιές μορφές είναι καλό να δημοσιοποιούνται τα δεδομένα, ώστε η περαιτέρω χρησιμοποίηση τους να είναι ευκολότερη (Schellong, 2011; Cyganiac, 2011).

Οι τρόποι και τα βήματα που πρέπει να ακολουθούνται από τις πρωτοβουλίες για την δημοσιοποίηση των ανοιχτών δημόσιων δεδομένων, περιγράφονται στο κεφάλαιο 2.5 (Bennett and Harvey, 2009) εξετάζοντας το σημαντικό θέμα των συνδεδεμένων δεδομένων και πως αυτό υλοποιείται στα κυβερνητικά δεδομένα (Berners-Lee, 2006, 2010; Bitzer et al., 2009). Οι κατάλογοι των δημόσιων ανοιχτών δεδομένων ως τρόπος για να φτάσουν αυτά πιο εύκολα στο ευρύ κοινό και η σύντομη παρουσίαση τους (Veljković, 2011; Kalampokis et al., 2011a) είναι η συνέχεια του κεφαλαίου και αυτό κλείνει με δύο παραδείγματα εφαρμογών που χρησιμοποιούν σύνολα δεδομένων τα οποία είναι ανοιχτά.

Το επόμενο κεφάλαιο ασχολείται με τα παρατηρημένα προβλήματα και τις προκλήσεις που πρέπει να ξεπεραστούν ώστε τα ανοιχτά δημόσια δεδομένα να εξυπηρετήσουν σωστά τον σκοπό για τον οποίο έγιναν ανοιχτά (Geiger and von Lucke, 2011; Peled, 2011, Audit Commission, 2011).

3.2. ΗΛΕΚΤΡΟΝΙΚΗ ΚΑΙ ΑΝΟΙΧΤΗ ΔΙΑΚΥΒΕΡΝΗΣΗ

3.2.1. ΕΙΣΑΓΩΓΗ

Η εξέλιξη και ανάπτυξη της τεχνολογίας των τηλεπικοινωνιών και των υπολογιστών προσέφερε στις κυβερνήσεις των χωρών ένα παραπάνω όπλο για να προσεγγίσουν τους πολίτες τους και να προσπαθήσουν να αλλάξουν τον τρόπο με τον οποίο φτάνουν οι υπηρεσίες τους στον κόσμο. Η υιοθέτηση των τεχνολογιών αυτών σηματοδότησε την Ηλεκτρονική Διακυβέρνηση η οποία πέρασε από διάφορα στάδια που καταγράφηκαν και μετράνε την πρόοδο των κυβερνήσεων όσον αφορά την ανάπτυξη σε σχέση με την Ηλεκτρονική Διακυβέρνηση. Η συγκέντρωση των υπηρεσιών σε μία ιστοσελίδα αποτελεί την ολοκλήρωση των παρεχομένων

υπηρεσιών. Η ανοιχτή διακυβέρνηση αποτελεί ένα νέο μοντέλο διακυβέρνησης, που αποτελεί εξέλιξη της Ηλεκτρονικής Διακυβέρνησης, και απαιτεί περισσότερη διαφάνεια, υπευθυνότητα και συνεργασία οδηγώντας στο άνοιγμα των κυβερνητικών δεδομένων.

3.2.2. ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΜΟΝΤΕΛΑ ΑΝΑΠΤΥΞΗΣ

Η Ηλεκτρονική Διακυβέρνηση, Η-Διακυβέρνηση, μπορεί να οριστεί ως η χρήση των Τεχνολογιών Πληροφορίας και Επικοινωνιών ΤΠΕ, από τις κυβερνήσεις ή κεντρικές διοικήσεις με σκοπό να ενισχύσουν το εύρος και την ποιότητα των παρεχόμενων πληροφοριών και υπηρεσιών προς τους πολίτες, τις επιχειρήσεις, τους οργανισμούς, τους φορείς τοπικής αυτοδιοίκησης αλλά και τους φορείς της κεντρικής διοίκησης με έναν αποτελεσματικό και οικονομικά προσφορότερο τρόπο καθιστώντας τις κυβερνητικές διαδικασίες πιο διαφανείς με απώτερο τελικό σκοπό την ενίσχυση της ίδιας της δημοκρατίας (Chen, 2009).

Σύμφωνα με τον Yildiz (2007) στις αρχές της ηλεκτρονικής διακυβέρνησης περιλαμβάνονται η διαφάνεια, η συμμετοχή και η συνεργασία. Η ανάπτυξη και η ενσωμάτωση της τεχνολογίας της πληροφορίας και της επικοινωνίας στην καθημερινότητα του πολίτη επιτρέπει να εφαρμοστούν αυτές οι τρεις αρχές στην πράξη. Η πεμπτουσία της διαφάνειας είναι η παροχή πληροφόρησης στους πολίτες. Στην ανοιχτή διακυβέρνηση το κράτος δεσμεύει όλες τις κρατικές υπηρεσίες και τους φορείς παροχής υπηρεσιών να παρέχουν όλες τις σχετικές πληροφορίες με τον κατάλληλο τρόπο. Σε μια ψηφιακή κοινωνία αυτό γίνεται εύκολα από τους ειδικούς των πληροφοριακών συστημάτων, καθώς τα δεδομένα και τα έγγραφα αποθηκεύονται σε μεγάλο βαθμό ηλεκτρονικά και μπορούν εύκολα να φορτωθούν σε ένα δικτυακό τόπο με το πάτημα ενός κουμπιού. Μέσω αυτών των ιστοσελίδων οι πολίτες μπορούν να έχουν πρόσβαση στα έγγραφα σχεδόν από οπουδήποτε χωρίς να χρειάζεται να επικοινωνήσουν με την υπηρεσία άμεσα, διατηρώντας έτσι τον φόρτο των κυβερνητικών υπηρεσιών στο ελάχιστο. Αυτή η ανοιχτή πληροφόρηση είναι η βάση για την ανοιχτή διακυβέρνηση, δεδομένου ότι μόνο η δωρεάν και εύκολη πρόσβαση σε σωστές πληροφορίες μπορεί να διευκολύνει τη συμμετοχή και τη συνεργασία. «*Η ΔΙΑΦΑΝΕΙΑ είναι η νέα αντικειμενικότητα. Η*

Διαφάνεια μας φέρνει την αξιοπιστία με τον τρόπο που η αντικειμενικότητα το έκανε»(Weinberger, 2009).

Οι έννοιες της διαφάνειας, της συμμετοχής και της συνεργασίας που αναδεικνύονται μέσω της ηλεκτρονικής διακυβέρνησης οδήγησαν σ` ένα άλλο επίπεδο τις υπηρεσίες των κυβερνήσεων προς τους πολίτες. Η εξέλιξη της αλληλεπίδρασης των πολιτών με το κράτος μέσω της τεχνολογίας των υπολογιστών και των τηλεπικοινωνιών δημιουργεί την ευκαιρία να προχωρήσει η έννοια της διακυβέρνησης από ηλεκτρονική σε ανοιχτή, την έννοια της οποίας θα εξηγήσουμε παρακάτω.

Ένα μοντέλο ανάπτυξης ηλεκτρονικής διακυβέρνησης παρουσιάστηκε σε μια μελέτη που διεξήχθη από τα Ηνωμένα Έθνη και την αμερικανική κοινωνία για τη Δημόσια Διοίκηση (UN and ASPA 2002, p. 2). Πρότεινε ένα μοντέλο με πέντε στάδια ανάπτυξης. Το πρώτο στάδιο είναι το «αναδυόμενο» στάδιο, στο οποίο υπάρχει μία επίσημη online παρουσία της κυβέρνησης. Στο δεύτερο στάδιο, ο αριθμός των ιστοσελίδων της κυβέρνησης αυξάνεται έτσι ώστε να γίνει πιο δυναμική σε αυτό το «ενισχυμένο» στάδιο. Το τρίτο «διαδραστικό» στάδιο επιτρέπει στους χρήστες να κατεβάσουν φόρμες και να αλληλεπιδρούν με τους υπαλλήλους μέσω του διαδικτύου. Στο τέταρτο στάδιο «συναλλαγής», οι χρήστες έχουν την δυνατότητα να κάνουν online πληρωμές για συναλλαγές. Το τέταρτο και τελικό «απρόσκοπτο» στάδιο καθιστά την ολοκλήρωση των ηλεκτρονικών υπηρεσιών σε όλους τους κυβερνητικούς οργανισμούς δυνατή. Το μοντέλο ASPA-UN είναι παρόμοιο με εκείνο του Layne and Lee (2001). Μετά το μοντέλο των Layne and Lee (2001) καταγράφηκαν μελέτες που βασίστηκαν σ` αυτό και επέκτειναν τα στάδια ωριμότητας της ηλεκτρονικής διακυβέρνησης. Ο Lee J. (2010) συγκέντρωσε τις κυριότερες από αυτές για να κάνει μία ποιοτική μετασύνθεση των μοντέλων ανάπτυξης την τελευταία δεκαετία προτείνοντας ένα κοινό πλαίσιο αναφοράς αποτελούμενο από δύο διαστάσεις, την διάσταση του πολίτη- υπηρεσίας και την διάσταση λειτουργίας-τεχνολογίας.

3.2.3. ONE-STOP E-GOVERNMENT

Ο όρος one-stop e-government portal σημαίνει ότι υπάρχει ένα κεντρικό σημείο (ιστοσελίδα) απ` όπου ο πολίτης μπορεί να πλοηγηθεί στις υπηρεσίες ή στα

δεδομένα που προσφέρονται από τις αρχές του τόπου του. Οι Dias and Rafael (2007) και ο Ταμπούρης (2001) ορίζουν την one-stop e-government ως την χρήση πληροφοριακών και τηλεπικοινωνιακών συστημάτων για την ολοκλήρωση των υπηρεσιών που προσφέρονται από το δημόσιο προς την πλευρά του πολίτη. Ο χρήστης θα μπορεί να έχει πρόσβαση στις υπηρεσίες είκοσι τέσσερις ώρες από όποια τοποθεσία έχει πρόσβαση στο διαδίκτυο. Επίσης οι μελέτες προτείνουν μία πλατφόρμα για να μπορεί να γίνει κατανοητή η έννοια της one-stop e-government.

Πολλές κυβερνητικές πρωτοβουλίες ηλεκτρονικής διακυβέρνησης επικεντρώνονται στην ανάπτυξη τέτοιων ιστοσελίδων και στην προσπάθεια της παρούσας εργασίας για την παρουσίαση των δημόσιων ανοιχτών δεδομένων, χρησιμοποιούμε τις παραπάνω έννοιες για να μπορέσουμε να κατανοήσουμε την παροχή των δεδομένων αυτών από ένα συγκεκριμένο σημείο (ιστοσελίδα).

3.2.4. ΑΠΟ ΤΗΝ ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΣΤΗΝ ΑΝΟΙΧΤΗ ΔΙΑΚΥΒΕΡΝΗΣΗ

Σύμφωνα με τους Alonso et al. (2009) οι κυβερνήσεις αγωνίζονται από τα τέλη της δεκαετίας του 1990 να βρούνε καλύτερους τρόπους για να επικοινωνήσουν με τους ψηφοφόρους τους μέσω του διαδικτύου. Από την στιγμή που τα κυβερνητικά δεδομένα γίνονται διαθέσιμα μέσω του internet, η εξεύρεση τους καθίσταται εύκολη, άμεσα διαθέσιμη, προσιτή, κατανοητή, τα δεδομένα μπορούν να χρησιμοποιηθούν και οι πολίτες μπορούν πλέον να αλληλεπιδρούν με την κυβέρνησή τους με τρόπους που ποτέ πριν δεν μπορούσε κάποιος να φανταστεί. Η έννοια αυτή ονομάστηκε «ηλεκτρονική διακυβέρνηση» ή e-Government.

Η υπόσχεση της Ηλεκτρονικής Διακυβέρνησης επιτρέπει στους πολίτες να έχουν πρόσβαση στις πληροφορίες και τα δεδομένα που παράγει η κυβέρνηση και οι υπηρεσίες της, με τους δικούς τους όρους. Η κοινή χρήση των δεδομένων της κυβέρνησης και των πληροφοριών της μαζί με το κοινό, προσφέρει διαφάνεια και βελτιώνει τις λειτουργίες εντός και μεταξύ των κυβερνήσεων.

Δυστυχώς, η αποτελεσματική ηλεκτρονική διακυβέρνηση δεν ήταν εύκολο να επιτευχθεί, λόγω των δυσκολιών που αντιμετωπίζουν οι κυβερνήσεις στη συλλογή, διαχείριση και διάθεση των πληροφοριών και των υπηρεσιών που διατίθενται ηλεκτρονικά. Οι προκλήσεις αυτές περιλαμβάνουν ξεπερασμένες πολιτικές,

δημοσιονομικούς και προσωπικούς περιορισμούς καθώς και αργή, γραφειοκρατική νοοτροπία. Το Web 2.0 και τα social media εμφανίστηκαν ως νέα παραδείγματα διαφάνειας και αλληλεπίδρασης προσθέτοντας νέες προκλήσεις στην προσαρμογή των κυβερνήσεων. Τα προβλήματα που έχουν να αντιμετωπίσουν οι κυβερνήσεις αυξήθηκαν και περιλαμβάνουν την διασφάλιση της αυθεντικότητας των πληροφοριών τους, όταν αυτές είναι ελεύθερα προσβάσιμες από το κοινό, την ένταξη των νέων τεχνολογιών στα ήδη υπάρχοντα συστήματα και την πρόσβαση σ` αυτές τις πληροφορίες από όλους τους πολίτες χωρίς διάκριση.

3.2.5. ΑΝΟΙΧΤΗ ΔΙΑΚΥΒΕΡΝΗΣΗ

Η εξέλιξη της ηλεκτρονικής διακυβέρνησης, περνώντας από τα στάδια που περιγράψαμε παραπάνω, μας οδηγεί σε ένα μοντέλο όπου οι βασικές αρχές της (διαφάνεια, συμμετοχή και συνεργασία) επιτυγχάνονται καθώς η διακυβέρνηση είναι πια ανοιχτή. Τι είναι όμως η ανοιχτή διακυβέρνηση; Στην πιο βασική της έννοια είναι η αντίληψη ότι οι άνθρωποι έχουν το δικαίωμα πρόσβασης στα έγγραφα και στις κυβερνητικές διαδικασίες. Η ιδέα ότι το κοινό έχει δικαίωμα να ελέγχει και να συμμετέχει σε διαδικασίες, τουλάχιστον για ενημέρωση, κατοχυρώνεται τόσο στη Διακήρυξη της Ανεξαρτησίας των ΗΠΑ αλλά και στο σύνταγμα. Οι αρχές της έχουν αναγνωριστεί σχεδόν σε κάθε δημοκρατική χώρα στον πλανήτη. Αλλά η ίδια η έννοια του όρου συνεχίζει να εξελίσσεται. Οι πολίτες δεν έχουν πρόσβαση μόνο στις πληροφορίες, τα έγγραφα και διαδικασίες, αλλά μπορούν επίσης να συμμετέχουν με έναν ουσιαστικό τρόπο. Ανοιχτή διακυβέρνηση σημαίνει επίσης βελτίωση της επικοινωνίας και των ενεργειών στο πλαίσιο των διαφόρων κλάδων και των επιπέδων της κυβέρνησης. Περισσότερη ανταλλαγή πληροφοριών στο εσωτερικό της μπορεί να οδηγήσει σε μεγαλύτερη αποτελεσματικότητα και υπευθυνότητα.

Ο Don Tapscott (2010) στην εισαγωγή του βιβλίου "Open Government" αναφέρει χαρακτηριστικά ότι το πρώτο κύμα των στρατηγικών της ηλεκτρονικής διακυβέρνησης έφερε μερικά σημαντικά οφέλη. Έκανε τις κυβερνητικές πληροφορίες και τις υπηρεσίες πιο προσιτές στους πολίτες, δημιουργώντας παράλληλα τη διοικητική και λειτουργική αποτελεσματικότητα. Αλλά πάρα πολλές από αυτές τις

πρωτοβουλίες επικεντρώθηκαν στην αυτοματοποίηση των υφιστάμενων διαδικασιών και στην προσφορά αυτών στο διαδίκτυο.

Είναι το επόμενο κύμα της καινοτομίας που παρουσιάζει μια ιστορική ευκαιρία να επανασχεδιαστεί ριζικά ο τρόπος λειτουργίας της κυβέρνησης, αυτά που ο δημόσιος τομέας παρέχει και τελικά τον τρόπο που αλληλεπιδρούν οι κυβερνήσεις και έρχονται σε επαφή με τους πολίτες τους. Είναι πραγματικά μια στιγμή που είτε η κυβέρνηση θα διαδραματίσει ενεργό και θετικό ρόλο στο μετασχηματισμό της, ή η αλλαγή απλά θα συμβεί. Η διαδικασία μετασχηματισμού είναι ταυτόχρονα συναρπαστική και επώδυνη, αλλά το κόστος της αδράνειας είναι μια χαμένη ευκαιρία για την κυβέρνηση να επαναπροσδιορίσει το ρόλο της στην κοινωνία και να βοηθήσει να ξεκινήσει μια νέα εποχή της συμμετοχικής διακυβέρνησης.

Συμπεραίνουμε λοιπόν ότι εμφανίζεται ένα νέο είδος οργάνωσης του δημόσιου τομέα και αυτό είναι η ανοιχτή διακυβέρνηση. Αυτός ο τρόπος διακυβέρνησης ανοίγει τις πόρτες της στον κόσμο, καινοτομεί μαζί με τους πολίτες, μοιράζεται τους πόρους που προηγουμένως ήταν καλά φυλαγμένοι, αξιοποιεί τη δύναμη της μαζικής συνεργασίας, έχει ως άξονα τη διαφάνεια σε όλες τις ενέργειές του και συμπεριφέρεται όχι ως ένα μεμονωμένο τμήμα, αλλά ως κάτι νέο, μια πραγματικά ολοκληρωμένη και δικτυωμένη οργάνωση.

3.3. ΔΗΜΟΣΙΑ ΑΝΟΙΧΤΑ ΔΕΔΟΜΕΝΑ

3.3.1. ΕΙΣΑΓΩΓΗ

Μετά την καταγραφή της εξέλιξης της ηλεκτρονικής διακυβέρνησης και την μετατροπή και προσαρμογή της στα σημερινά δεδομένα, είδαμε την εμφάνιση ενός νέου είδους οργάνωσης του δημόσιου τομέα που ονομάζεται ανοιχτή διακυβέρνηση. Η βάση αυτού του νέου τρόπου οργάνωσης είναι τα δημόσια ανοιχτά δεδομένα των οποίων η παραγωγή με τις καινούργιες τεχνολογίες των υπολογιστών γίνεται όλο και μεγαλύτερη σε όγκο καθιστώντας τον δημόσιο τομέα (Public Sector) τον μεγαλύτερο παραγωγό διαφόρων τύπων δεδομένων. Για να μπορέσουμε να προχωρήσουμε στην ανάλυση τους θα πρέπει πρώτα να ορίσουμε τι

είναι δεδομένα, τι είναι δημόσια δεδομένα, τι είναι ανοιχτά δεδομένα και τέλος τι είναι δημόσια ανοιχτά δεδομένα. Πριν από αυτό όμως θα κάνουμε μια αναδρομή και θα εξετάσουμε την αντιμετώπιση της ανοιχτής διακυβέρνησης από τις ΗΠΑ και την Ευρωπαϊκή Ένωση μέσω των νόμων που έχουν εκδώσει κατά καιρούς και αφορούν τα δημόσια ανοιχτά δεδομένα (Public Sector Information, PSI).

3.3.2. Η ΚΟΙΝΩΝΙΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

Οι ΗΠΑ πρωτοπορώντας στην ανάπτυξη και την εφαρμογή των ανοιχτών στρατηγικών της κυβέρνησης ξεκίνησαν τις συζητήσεις με την χρήση του όρου «ανοιχτή διακυβέρνηση» από τα τέλη της δεκαετίας του 1950, λόγω του γεγονότος ότι μετά τον Δεύτερο Παγκόσμιο Πόλεμο οι δημοσιεύσεις καθυστερούσαν όλο και περισσότερο και μερικές φορές οι πληροφορίες παρακρατούνταν. Οι αρχές της ανοικτής διακυβέρνησης και η ιδέα της ελεύθερης πληροφόρησης από την κυβέρνηση άρχισαν να εξετάζονται. Ως αποτέλεσμα, νομικές τροποποιήσεις έλαβαν χώρα οδηγώντας στον νόμο περί ελευθερίας της πληροφορίας (Freedom of Information Act, FOIA), ο οποίος καθορίστηκε το 1966 και τέθηκε σε ισχύ το 1967 παρά την αντίθεση του Προέδρου Johnson εκείνη την εποχή. Ο συγκεκριμένος νόμος δίνει στους Αμερικανούς πολίτες το δικαίωμα να ζητήσουν επίσημα έγγραφα. Η διατύπωση των διαφόρων τροποποιήσεων του FOIA καθιστά σαφές ότι το κράτος υποχρεούται να αποκαλύψει όλες τις πληροφορίες, ως εκ τούτου χρειάζεται να δικαιολογήσει την απαγόρευση της πρόσβασης. Ταξινομημένα δεδομένα, όπως εμπιστευτικά στρατιωτικά και οικονομικά στοιχεία, η εσωτερική επικοινωνία στα πλαίσια των αρχών και ιδιωτικά δεδομένα των πολιτών πρέπει ή δεν χρειάζεται να αποκαλυφθούν. Σε περίπτωση που μία κυβερνητική υπηρεσία δεν πληροί την υποχρέωση παροχής πληροφοριών, θα ακολουθήσουν κυρώσεις (Parycek and Sachs, 2010).

Η τροποποίηση του νόμου (FOIA) το 1996 κατέστησε σαφές ότι οι αμερικανικές αρχές πρέπει να δώσουν έμφαση στην πρόσβαση στις πληροφορίες μέσω υπολογιστών. Οι αρχές υποχρεούνται να χορηγούν στους πολίτες πρόσβαση σε τεχνικά μέσα, εάν είναι απαραίτητο. Λόγω της ραγδαίας ανάπτυξης της τεχνολογίας της πληροφορίας και της επικοινωνίας (ΤΠΕ) το FOIA τροποποιήθηκε

και πάλι το 2007 με «the Openness Promotes Effectiveness in our National Government Act» του 2007.

Την πρώτη μέρα μετά την ορκωμοσία του, ο πρόεδρος Ομπάμα δήλωσε ότι θα αναπτύξει περαιτέρω τις αρχές της ανοικτής διακυβέρνησης στην κυβέρνησή του. «*Θα εργαστούμε από κοινού για να εξασφαλιστεί η εμπιστοσύνη του κοινού και να δημιουργηθεί ένα σύστημα διαφάνειας, συμμετοχής του κοινού, και συνεργασίας. Η διαφάνεια θα ενισχύσει τη δημοκρατία μας και την προώθηση της αποδοτικότητας και της αποτελεσματικότητας στην κυβέρνηση*» (Obama, 2009). Το αποτέλεσμα του Open Government Memorandum του προέδρου Ομπάμα είναι η Open Government Directive (Orszag, 2009) που υποχρεώνει τις αρχές να δημοσιεύουν τα στοιχεία και τα έγγραφα μέσα σε ένα ορισμένο χρονικό πλαίσιο. Τα δεδομένα αυτά πρέπει να είναι σύμφωνα με τις κατευθυντήριες γραμμές που είναι η διασφάλιση και μεγιστοποίηση της ποιότητας, η αντικειμενικότητα, η χρησιμότητα και η ακεραιότητα των πληροφοριών (συμπεριλαμβανομένων των στατιστικών στοιχείων) που δίδονται από τον οργανισμό. Όλες οι κυβερνητικές υπηρεσίες υποχρεούνται να αναπτύξουν στρατηγικές για την εφαρμογή της ανοικτής διακυβέρνησης στους αντίστοιχους τομείς σε αρμονία με την τελευταία τεχνολογία για μέγιστο όφελος. Ο στόχος είναι να βελτιωθεί η ποιότητα της ανοικτής διακυβέρνησης, η δημιουργία μίας κουλτούρας ανοικτής διακυβέρνησης, και να εκπονηθούν κατευθυντήριες γραμμές για την ανοιχτή διακυβέρνηση στις κυβερνητικές υπηρεσίες.

Το δικαίωμα στην πληροφορία έχει και μια μακρά και μια πολύ πρόσφατη ιστορία. Αν και ο πρώτος νόμος ανάγεται στο 1776 και στο Σουηδικό νόμο για την Ελευθερία του Τύπου, η πλειονότητα των Ευρωπαϊκών χωρών υιοθέτησαν νόμους για Freedom Of Information κατά τα τελευταία 15 χρόνια. Το 1990 υπήρχαν μόλις 12 νόμοι, τώρα υπάρχουν άνω των 65. Οι πρώιμοι Ευρωπαϊκοί νόμοι (Σουηδία , Φιλανδία , Νορβηγία , Γαλλία και Ολλανδία – όλοι πριν το 1980) και ο νόμος των Η.Π.Α για την Ελευθερία στην πληροφόρηση (FOIA) κωδικοποίησαν τις διοικητικές διαδικασίες για την – μετά από αίτηση - παροχή πληροφοριών στο κοινό δίνοντας έμφαση σε έγγραφα της διοίκησης (και ελάχιστα έως καθόλου σε νομοθετικά, εκτελεστικά ή δικαστικά έγγραφα). Έτσι οι πρώτοι αυτοί νόμοι καθιέρωσαν ένα επίσημο δικαίωμα για την πρόσβαση σε διοικητικά έγγραφα ενώ ελάχιστοι έκαναν

αναφορά στο γενικότερο δικαίωμα πρόσβασης στην πληροφορία (Access to information, 2006).

3.3.3. ΟΡΙΣΜΟΙ

Δεδομένα: Τα δεδομένα, ο αγγλικός όρος data από τον πληθυντικό του λατινικού datum, είναι ένα σύνολο διακριτών, αντικειμενικών στοιχείων για γεγονότα. Μπορεί να είναι αριθμοί, λέξεις, σύμβολα, γεγονότα, που περιγράφουν ή αντιπροσωπεύουν ποσότητες, έννοιες, ιδέες, αντικείμενα, καταστάσεις και λειτουργίες. Στην πληροφορική συναντούμε τα δεδομένα στον πληθυντικό αριθμό, σπανιότερα στον ενικό (δεδομένο). Επίσης, δεδομένο ονομάζεται ένα γνωστό ή αποδεκτό στοιχείο το οποίο χρησιμοποιείται ως βάση ή προϋπόθεση στην επίλυση προβλημάτων. Τα δεδομένα περιγράφουν μόνο μέρος ενός συμβάντος, δεν περιλαμβάνουν καμία ανάλυση, κριτική ή αξιόπιστη βάση για περαιτέρω ενέργεια. Τα δεδομένα δεν αναφέρουν τίποτα για τη σημαντικότητά τους ή τη σχέση τους προς οτιδήποτε. Σε αντίθεση με τα δεδομένα η πληροφορία έχει ορισμένο νόημα και είναι οργανωμένη για συγκεκριμένο σκοπό. Με την προσθήκη αξιών στα δεδομένα με διαμόρφωση, οργάνωση, μαθηματική ή στατιστική ανάλυση, διόρθωση λαθών ή συμπίεση παράγεται η πληροφορία. Τη διαδικασία αυτή ονομάζουμε επεξεργασία δεδομένων (Davenport and Prusak, 1998).

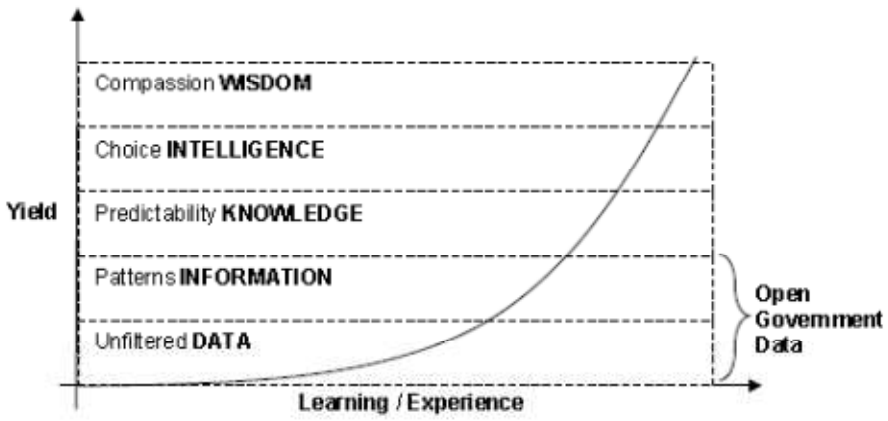
Η βιβλιογραφία της διαχείρισης της γνώσης παρέχει μια πλούσια συζήτηση για θέματα που σχετίζονται με τα δεδομένα, πληροφορίες, γνώσεις και μεταξύ τους σχέσεις. Η συμβατική άποψη είναι ότι υπάρχει μια ιεραρχική σχέση μεταξύ δεδομένων, πληροφοριών και γνώσεων (εικόνα 3.1). Καθώς τα δεδομένα και τα αντικείμενα πληροφοριών (π.χ. μια αναφορά) μπορούν να μεταφέρουν τη δικής τους αναπαράσταση της πληροφορίας, είναι ο συνδυασμός του περιεχομένου και της ερμηνείας που οι χρήστες βρίσκουν πολύτιμο. Αυτή είναι η ιδέα της καμπύλης στην εικόνα. Καθώς τα δεδομένα βελτιώνονται όλο και περισσότερο, η απόδοση των διαφόρων μορφών των δεδομένων, πληροφοριών και γνώσεων για ένα άτομο ή την κοινωνία αυξάνει. Αυτή η συζήτηση μπορεί να είναι της φιλοσοφικής φύσης, αλλά είναι απαραίτητη γιατί τα δημόσια ανοιχτά δεδομένα παρέχουν σπάνια καθοδήγηση σχετικά με την έννοια των δεδομένων. Οι όροι συχνά

χρησιμοποιούνται περιστασιακά γεγονός το οποίο οδηγεί σε σύγχυση εντός και εκτός της κυβέρνησης, π.χ. κατά τη συζήτηση της έννοιας των πρωτογενών δεδομένων. Το ίδιο ισχύει και για το Public Sector Information (PSI) που χρησιμοποιείται συχνά ως ένας γενικός όρος για το σύνολο του περιεχομένου που παράγεται από δημόσιους φορείς.

Ανοιχτά δεδομένα: Ο ορισμός των ανοιχτών δεδομένων σύμφωνα με το Open Knowledge Foundation (Open Definition, 2011a) μπορεί περιληπτικά να συνοψισθεί στα παρακάτω :

«*Ανοιχτό είναι ένα κομμάτι περιεχομένου ή δεδομένων εφόσον οποιοσδήποτε είναι ελεύθερος να το χρησιμοποιήσει, επαναχρησιμοποιήσει και επαναδιαθέσει, χωρίς περιορισμούς ή με μόνο περιορισμό την αναφορά στον δημιουργό*».

Ο πλήρης ορισμός (Open Definition, 2011b) θέτει έντεκα όρους για να μπορούν τα δεδομένα να θεωρηθούν ανοιχτά.



Εικόνα 3.1 Δεδομένα με τη συμβατική άποψη σχετικά με την ιεραρχία της γνώσης (Schellong, 2011)

Δημόσια δεδομένα: Όταν αναφερόμαστε στα δημόσια ή κυβερνητικά δεδομένα εννοούμε δεδομένα και πληροφορίες που παράγονται από την κυβέρνηση ή κυβερνητικούς φορείς καθώς και αυτά που έχει ανατεθεί η παραγωγή τους σε οντότητες που ελέγχονται από την κυβέρνηση. Συμπερασματικά από τον ορισμό καταλαβαίνουμε ότι δημόσια δεδομένα μπορεί να παράγει και ο ιδιωτικός τομέας ή

φορείς αυτού αρκεί αυτά να σχετίζονται με το δημόσιο (Open Government Data 2011).

Με τον όρο «Δημόσια Δεδομένα» εννοούμε δεδομένα και πληροφορίες που σχετίζονται με την δημόσια σφαίρα ανεξάρτητα του εάν παράγονται από δημόσιους ή ιδιωτικούς φορείς. Ο όρος δημόσια δεδομένα λοιπόν χρησιμοποιείται με την διασταλτική έννοια του όρου και μπορεί να περιλαμβάνει δεδομένα που :

- Βρίσκονται στην κατοχή ενός οργανισμού του στενού ή ευρύτερου δημόσιου τομέα.
- Έχουν παραχθεί με δημόσιους πόρους.
- Παράγονται από τον ιδιωτικό τομέα αλλά αφορούν γενικότερα το κοινωνικό σύνολο.

Σύμφωνα με τα παραπάνω, στα δημόσια δεδομένα δεν περιλαμβάνονται αποκλειστικά δεδομένα που σχετίζονται με κρατικές , διοικητικές ή πολιτειακές αρμοδιότητες. Μπορεί να συμπεριλαμβάνονται ακόμα και δεδομένα που παράγονται από ιδιωτικούς φορείς και επιχειρήσεις, οι οποίοι έχουν χρηματοδοτηθεί για την παραγωγή τους από κρατικά κονδύλια. Ακόμα όμως και στις περιπτώσεις που δεν υφίσταται κρατική χρηματοδότηση , τα δεδομένα που παράγει ο Ιδιωτικός τομέας μπορεί να θεωρηθούν δημόσια αρκεί να σχετίζονται με την δημόσια σφαίρα (Παπαδόπουλος, 2011). Ποιοτικά, δημόσια δεδομένα παράγει δυνητικά ο κάθε πολίτης. Πρώτα από όλα, όμως, το κάνει το κράτος στο όνομά τους. Το υποσύνολο των δημοσίων δεδομένων που παράγονται από το κράτος αποκαλούνται Κυβερνητικά Δεδομένα.

Παράδειγμα δημοσίων δεδομένων είναι τα δεδομένα που αφορούν τα σήματα του μηχανισμού της αγοράς ή τα δεδομένα σχετικά με τον καιρό. Στα δημόσια δεδομένα όμως συμπεριλαμβάνονται όλα τα κυβερνητικά δεδομένα, γιατί αφορούν την σφαίρα του συλλογικού. Μάλιστα πολλές φορές το κράτος αναλαμβάνει την παραγωγή δεδομένων που είναι καθαυτά δημόσια και δεν σχετίζονται με κυβερνητικές αρμοδιότητες. Υπάρχουν αρκετοί λόγοι για τους οποίους το Κράτος παράγει δημόσια –μη κυβερνητικά- δεδομένα. Ο βασικός είναι γιατί η αγορά δεν μπορεί να αναλάβει τη παραγωγή τους. Είτε γιατί το κόστος παραγωγής είναι πολύ υψηλό, είτε γιατί τα συγκεκριμένα δεδομένα δεν μπορούν να

διατεθούν ως προϊόν από την αγορά. Τέτοια παραδείγματα αποτελούν οι δορυφορικές απεικονίσεις ή οι αεροφωτογραφίες το κόστος παραγωγής των οποίων είναι ιδιαίτερα υψηλό ή τα στοιχεία σχετικά με τη ποιότητα των υδάτων και τους ατμοσφαιρικούς ρύπους τα οποία δεν θεωρούνται αξιοποιήσιμα από την αγορά.

Δημόσια ανοιχτά δεδομένα: Χρησιμοποιώντας τον ορισμό του Open Knowledge Foundation για τα ανοιχτά δεδομένα και συγχωνεύοντας τον με αυτόν για τα δημόσια δεδομένα μπορούμε να ορίσουμε τα δημόσια ανοιχτά δεδομένα ως αυτά που είναι ανοιχτά και παράγονται από την κυβέρνηση ή κυβερνητικούς φορείς καθώς και αυτά που έχει ανατεθεί η παραγωγή τους σε οντότητες που ελέγχονται από την κυβέρνηση και μπορούν ελεύθερα να χρησιμοποιηθούν, να επαναχρησιμοποιηθούν και να επαναδιατεθούν από τον καθένα (Open Government Data, 2011). Μια γραφική απεικόνισή τους φαίνεται στην εικόνα 3.2.

Εικόνα 3.2 Το οικοσύστημα των ανοιχτών δεδομένα (Παπαδόπουλος, 2011)

3.3.4. ΑΡΧΕΣ ΤΩΝ ΑΝΟΙΧΤΩΝ ΚΥΒΕΡΝΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Στις 8 Δεκεμβρίου 2007, τριάντα υποστηρικτές της ανοιχτής διακυβέρνησης συγκεντρώθηκαν για να αναπτύξουν ένα σύνολο αρχών για τα ανοικτά δημόσια δεδομένα. Η συνάντηση, που πραγματοποιήθηκε στη Σεβαστούπολη της

Καλιφόρνια, είχε ως σκοπό να καταδείξει γιατί τα ανοιχτά κυβερνητικά δεδομένα είναι απαραίτητα για τη δημοκρατία.

Το Διαδίκτυο είναι ο δημόσιος χώρος του σύγχρονου κόσμου και μέσω αυτού οι κυβερνήσεις έχουν τώρα την ευκαιρία να κατανοήσουν καλύτερα τις ανάγκες των πολιτών και οι πολίτες μπορούν να συμμετέχουν πλήρως στην κυβέρνησή τους. Τα ανοιχτά δεδομένα προωθούν την ενισχυμένη πολιτική συζήτηση, τη βελτίωση της ευημερίας των πολιτών, καθώς και μια πιο αποτελεσματική χρήση των δημόσιων πόρων. Η ομάδα εργασίας προσφέρει ένα σύνολο θεμελιωδών αρχών για τα ανοικτά κυβερνητικά δεδομένα. Εφαρμόζοντας τις οκτώ αρχές, οι κυβερνήσεις του κόσμου μπορούν να γίνουν πιο αποτελεσματικές, διαφανής, και σχετικότερες με τη ζωή μας.

Τη συνάντηση συντόνισε ο Tim O`Reilly της O`Reilly Media και ο Carl Malamud της Public.Resource.Org με την χορηγία του Sunlight Foundation της Google και της Yahoo (8 Principles of Open Government Data, 2007) και το αποτέλεσμα αυτής της ομάδας εργασίας είναι οι παρακάτω οκτώ αρχές που πρέπει να χαρακτηρίζουν τα κυβερνητικά δεδομένα για να θεωρηθούν ανοιχτά:

1. **Πληρότητα:** Τα δεδομένα που διατίθενται πρέπει να είναι όσο το δυνατόν πιο πλήρη. Όλη η πρωτογενής πληροφορία για ένα σύνολο δεδομένων πρέπει να είναι προσβάσιμη από όλους εκτός από δεδομένα που αφορούν προσωπικά δεδομένα
2. **Πρωτογενή δεδομένα:** Τα δεδομένα πρέπει να διατίθενται όπως συλλέγονται από την πηγή τους με την μεγαλύτερη δυνατή αναλυτικότητα, όχι σε συνδυασμένες ή τροποποιημένες μορφές.
3. **Αμεσότητα:** Τα δεδομένα πρέπει να διατίθενται όσο πιο γρήγορα γίνεται στο κοινό, για να διατηρηθεί η αξία των δεδομένων. Τα δεδομένα που είναι χρονικά ευαίσθητα πρέπει να διατίθενται άμεσα, ακόμη αν είναι δυνατόν και σε πραγματικό χρόνο έτσι ώστε η πληροφορία να μην είναι αλλοιωμένη.
4. **Προσβασιμότητα:** Τα δεδομένα πρέπει να είναι διαθέσιμα και να μπορούν να έχουν πρόσβαση σ` αυτά ηλεκτρονικά το μεγαλύτερο κομμάτι των χρηστών για οποιονδήποτε λόγο. Τα δεδομένα πρέπει να είναι διαθέσιμα στο Διαδίκτυο ώστε να μπορούν να εξυπηρετήσουν το μεγαλύτερο φάσμα

χρηστών και χρήσεων. Αυτό σημαίνει ότι πρέπει να λαμβάνεται υπ` όψιν πως οι επιλογές στην προετοιμασία και την διάθεση των δεδομένων επηρεάζουν την πρόσβαση στα άτομα με ειδικές ανάγκες και τους χρήστες διαφορετικών λογισμικών και τεχνολογιών. Τα δεδομένα πρέπει να διατίθενται σύμφωνα με τα τρέχοντα πρωτόκολλα και πρότυπα καθώς και με εναλλακτικά όταν η τεχνολογία προβάλλει εμπόδια στην ευρεία επαναχρησιμοποίηση των δεδομένων.

5. **Αναγνωσιμότητα από υπολογιστές:** Τα δεδομένα πρέπει να είναι λογικά δομημένα ώστε να επιτρέπεται η αυτοματοποιημένη επεξεργασία τους. Για να υπάρχει η δυνατότητα ευρείας επαναχρησιμοποίησης των δεδομένων, πρέπει αυτά να είναι κατάλληλα κωδικοποιημένα. Το ελεύθερο κείμενο δεν είναι υποκατάστατο των πινάκων ή των κανονικοποιημένων πεδίων, καθώς και οι εικόνες του κειμένου. Επαρκής τεκμηρίωση (μεταδεδομένα) για την μορφή των δεδομένων και το νόημα των κανονικοποιημένων δεδομένων πρέπει να είναι διαθέσιμα στον χρήστη. Λαμβάνοντας υπόψη τη αρχή της προσβασιμότητας των δεδομένων, αυτή πρέπει να επεκτείνεται στην αυτοματοποιημένη πρόσβαση. Αν τα δεδομένα είναι προσβάσιμα από κάποια διεπαφή, πρέπει να είναι δυνατό το πλήρες σύνολο των δεδομένων να «κατέβει» τοπικά στον υπολογιστή σε πρωτογενή μορφή μέσω μιας αυτοματοποιημένης διαδικασίας.
6. **Προσβασιμότητα χωρίς διακρίσεις:** Τα δεδομένα πρέπει να είναι διαθέσιμα στον καθένα χωρίς την απαίτηση κάποιας εγγραφής. Η ανώνυμη πρόσβαση στα δεδομένα πρέπει να επιτρέπεται για τα δημόσια δεδομένα και δεν πρέπει αυτά να κρύβονται πίσω από «περιφραγμένους κήπους» και να είναι προσβάσιμα μόνο από συγκεκριμένες κατηγορίες χρηστών του διαδικτύου ή μέσω συγκεκριμένων εφαρμογών.
7. **Μη αποκλειστικές μορφές δεδομένων (Ανοιχτά Πρότυπα):** Τα δεδομένα πρέπει να είναι διαθέσιμα σε μορφές πάνω στις οποίες καμία οντότητα δεν έχει αποκλειστικό έλεγχο. Οι αποκλειστικές μορφές δεδομένων προσθέτουν περιττούς περιορισμούς στην χρήση, την κοινή χρήση και την επαναχρησιμοποίηση των δεδομένων. Παρόλο που μερικές μορφές «κλειστών προτύπων» τις συναντάμε σχεδόν παντού, δεν είναι σωστό να

γίνεται αποκλειστική χρήση τους. Φυσικά και κάποιες μορφές ανοικτών προτύπων δεν θα μπορέσουν να φτάσουν σε μεγάλο κομμάτι των χρηστών. Σ` αυτές τις περιπτώσεις προτείνεται τα δεδομένα να διατίθενται σε πολλαπλές μορφές.

8. **Χωρίς Αδειοδότηση:** Τα δεδομένα δεν πρέπει να υπόκεινται σε κανένα όρο χρήσης ή μυστικό κανονισμό. Η προστασία της ιδιωτικής ζωής, της ασφάλειας και περιορισμοί προνομίων επιτρέπονται καθώς διέπονται από άλλη νομολογία. Επειδή οι κυβερνητικές πληροφορίες είναι ένα μείγμα από δημόσια αρχεία, προσωπικές πληροφορίες, πνευματική ιδιοκτησία και άλλα μη δημόσια δεδομένα, είναι σημαντικό να υπάρχει σαφήνεια για το τι δεδομένα είναι διαθέσιμα και αν υπάρχουν άδεις, όροι χρήσης και νομικοί περιορισμοί που επιβάλλονται σ` αυτά. Δεδομένα για το οποία δεν υπάρχουν περιορισμοί θα πρέπει να διατίθενται στο κοινό (Annotations to the 8 Principles, 2007).

Οι παραπάνω αρχές εξετάζουν σε πολύ μεγάλο βαθμό όλες τις πλευρές των δημόσιων κυβερνητικών δεδομένων που θα πρέπει κάποιος να λάβει υπόψη του για να ξεκινήσει την διάθεση συνόλων δεδομένων στο διαδίκτυο για περαιτέρω χρήση και επαναχρησιμοποίηση. Ο David Eaves (2009) από μία δική του οπτική γωνία και χωρίς να ξεφεύγει από τα πλαίσια των αρχών του OKF συνοψίζει τις αρχές σε τρεις Νόμους των Δημόσιων Κυβερνητικών Δεδομένων:

- Αν τα δεδομένα δεν μπορούν να εντοπιστούν από τις μηχανές αναζήτησης ή να ευρετηριοποιηθούν δεν υπάρχουν.
- Αν δεν είναι διαθέσιμα σε ανοιχτή και αναγνώσιμη από υπολογιστή μορφή, δεν μπορούν να χρησιμοποιηθούν.
- Αν το νομικό πλαίσιο δεν επιτρέπει την επεξεργασία τους, τότε δεν υπάρχει αξία.

«*Βρες, Χρησιμοποίησε και Μοιράσου*» είναι τα λόγια που χρησιμοποιεί ο ίδιος, για να εξηγήσει τους νόμους που διατυπώνει. Αν οι μηχανές αναζήτησης δεν μπορούν να εντοπίσουν τα δεδομένα τότε πρακτικά δεν υπάρχουν για τους περισσότερους. Αφού το σύνολο δεδομένων φτάσει στον υπολογιστή του χρήστη πρέπει αυτός να

μπορεί να χρησιμοποιήσει τα δεδομένα, δηλαδή να τα συνδυάσει με άλλα δεδομένα, να πάρει αποτελέσματα από πίνακες, να τα μετατρέψει σε κάποια μορφή ή να χρησιμοποιήσει κάποιο πρόγραμμα της επιλογής του. Τέλος αφού ο χρήστης βρει και χρησιμοποιήσει τα δεδομένα, χρειάζεται ένα νομικό πλαίσιο που να του επιτρέπει να μοιραστεί αυτό που έχει δημιουργήσει, να είναι σε θέση να κινητοποιήσει τους υπόλοιπους πολίτες, να παρέχει νέες υπηρεσίες ή απλά να επισημάνει ένα ενδιαφέρον γεγονός. Αυτό σημαίνει ότι οι πληροφορίες και δεδομένα πρέπει να έχουν άδεια που να επιτρέπει την πιο ελεύθερη πιθανή χρήση ή, στην ιδανική περίπτωση, να μην έχουν κανέναν περιορισμό. Τα καλύτερα στοιχεία και πληροφορίες της κυβέρνησης, είναι αυτά που δεν προστατεύονται από πνευματικά δικαιώματα.

3.3.5. ΧΡΗΣΙΜΟΤΗΤΑ ΔΗΜΟΣΙΟΠΟΙΗΣΗΣ ΤΩΝ ΑΝΟΙΧΤΩΝ ΚΥΒΕΡΝΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Το μεγαλύτερο μέρος της βιβλιογραφίας που ασχολείται με το θέμα των ανοιχτών κυβερνητικών δεδομένων και της ανοιχτής διακυβέρνησης συγκλίνουν στην άποψη ότι κάνοντας οι κυβερνήσεις τα δεδομένα που παράγουν ανοιχτά στο κοινό, βοηθούν στην βελτίωση της διαφάνειας προάγουν τη συμμετοχή των πολιτών στα κοινά, φέρνουν μεγαλύτερη υπευθυνότητα και από τις δύο πλευρές και ενισχύουν την καινοτομία (Schellong and Stepanets, 2011; Parycek and Sachs, 2010; Maali et al., 2010; Lathrop and Ruma, 2010; Alonso et al., 2009; Linders and Wilson, 2011; Tauberer, 2009; White House, 2011).

Ο Tim O`Reilly (2010) διευκρινίζει το τι σημαίνει για τον ίδιο συμμετοχή, στο περιβάλλον όπου τα δεδομένα των κυβερνήσεων είναι ανοιχτά και ο πολίτης μπορεί να έχει ενεργή συμμετοχή στην λήψη των αποφάσεων μέσω εφαρμογών που είναι σχεδιασμένες πάνω σε δημόσια δεδομένα και όχι απλά να δίνει την άποψη του υπό μορφή σχολίων. Επίσης τονίζει την δημιουργία καινοτομίας από τους πολίτες και τις επιχειρήσεις όταν οι κυβερνήσεις χρησιμοποιούν ανοιχτά πρότυπα και διευκολύνουν την χρήση των δεδομένων από αυτούς που κατασκευάζουν καινοτόμες εφαρμογές. Στην ίδια λογική κινείται και η αναφορά του εγγράφου της κυβέρνησης του Ηνωμένου Βασιλείου στην καινοτομία και στις ευκαιρίες που παρουσιάζονται με την χρήση των Ανοιχτών Κυβερνητικών Δεδομένων. (HM Government, 2009) Την

συνεργασία των U.S. Department of Transport, Environmental Protection Agency και Housing and Urban Development με κυβερνητικές ομάδες, πολίτες και ανθρώπους της τεχνολογίας ώστε να βρουν νέες πρακτικές και καινοτόμες χρήσεις για τα δημόσια δεδομένα μας ενημερώνει ο Λευκός Οίκος (White House, 2011).

Ο Tim Berners-Lee (2009) σημειώνει ότι κυβερνητικά δεδομένα πρέπει να είναι διαθέσιμα στο διαδίκτυο για τους παρακάτω λόγους:

- Την αύξηση της επίγνωσης των πολιτών για τις κυβερνητικές διεργασίες έτσι ώστε να ενεργοποιήσει μεγαλύτερη υπευθυνότητα.
- Να συνεισφέρει πολύτιμες πληροφορίες για τον κόσμο.
- Την ενεργοποίηση της κυβέρνησης, της χώρας και του κόσμου ώστε να λειτουργήσουν πιο αποτελεσματικά.

Ο Gray J. (2011) αναφέρει, στην παρουσίαση που έκανε στην Ρώμη στις 19/4/2011, για την χρησιμότητα των Ανοιχτών Κυβερνητικών Δεδομένων τα παρακάτω:

- Καλύτερες υπηρεσίες από το δημόσιο τομέα.
- Οικονομική ανάπτυξη.
- Υπευθυνότητα.
- Καλύτερη διακυβέρνηση.
- Καινοτομία.
- Αποτελεσματικότητα.

Μελετώντας λοιπόν βλέπουμε ότι οι συγγραφείς που ασχολούνται με το θέμα των ανοιχτών δεδομένων σε γενικές γραμμές συμφωνούν στους λόγους για τους οποίους τα κυβερνητικά δεδομένα πρέπει να δημοσιεύονται και να είναι διαθέσιμα σε όποιον θέλει να τα χρησιμοποιήσει. Δεν θα πρέπει να οδηγηθούμε από το πνεύμα απόδοσης ευθυνών για τα λάθη της δημόσιας διοίκησης, αλλά περισσότερο στο χτίσιμο μιας συνεργασίας με εμπιστοσύνη μεταξύ των πολιτών και της κυβέρνησης όπου θα υπάρχουν προτάσεις για λύσεις, θα αμβλυνθεί η αναποτελεσματικότητα και θα υπάρξει βελτίωση στις υπηρεσίες από το κράτος στο πολίτη. Με αυτόν τον τρόπο οι κυβερνήσεις θα βελτιώσουν τον τρόπο με τον οποίο λαμβάνουν αποφάσεις και το πολιτικό σύστημα θα γίνει πιο δίκαιο.

3.4. ΕΠΑΝΑΧΡΗΣΙΜΟΠΟΙΗΣΗ ΤΩΝ ΑΝΟΙΧΤΩΝ ΚΥΒΕΡΝΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

3.4.1. ΕΙΣΑΓΩΓΗ

Ο λόγος που τα κυβερνητικά δεδομένα γίνονται ανοιχτά προς τους πολίτες είναι κυρίως η διαφάνεια και η υπευθυνότητα των δημοσίων υπηρεσιών. Είναι όμως και η δημιουργία καινοτομίας και προστιθέμενης αξίας που δημιουργείται από την περαιτέρω χρησιμοποίηση τους. Για αυτό το λόγο η Ευρωπαϊκή Ένωση αλλά και η Αμερικάνικη κυβέρνηση θεσμοθέτησαν συγκεκριμένο νομοθετικό πλαίσιο ώστε να βοηθήσουν και να προάγουν την επαναχρησιμοποίηση τους. Η μορφή των δεδομένων παίζει σημαντικό ρόλο στο να μπορέσει κάποιος να τα «κατεβάσει» στον υπολογιστή του και να προσθέσει εμπορική αξία σ` αυτά μέσω της επεξεργασίας τους. Οι κατηγορίες των δεδομένων είναι πολλές, ξεκινώντας από γεωχωρικά και καταλήγοντας σε οικονομικά και πολιτιστικά, και πάνω σ` αυτά υπάρχει η δυνατότητα να χτιστούν διαφόρων ειδών εφαρμογές.

3.4.2. ΕΥΡΩΠΑΪΚΗ ΕΝΩΣΗ ΚΑΙ ΝΟΜΙΚΟ ΠΛΑΙΣΙΟ ΕΠΑΝΑΧΡΗΣΙΜΟΠΟΙΗΣΗΣ

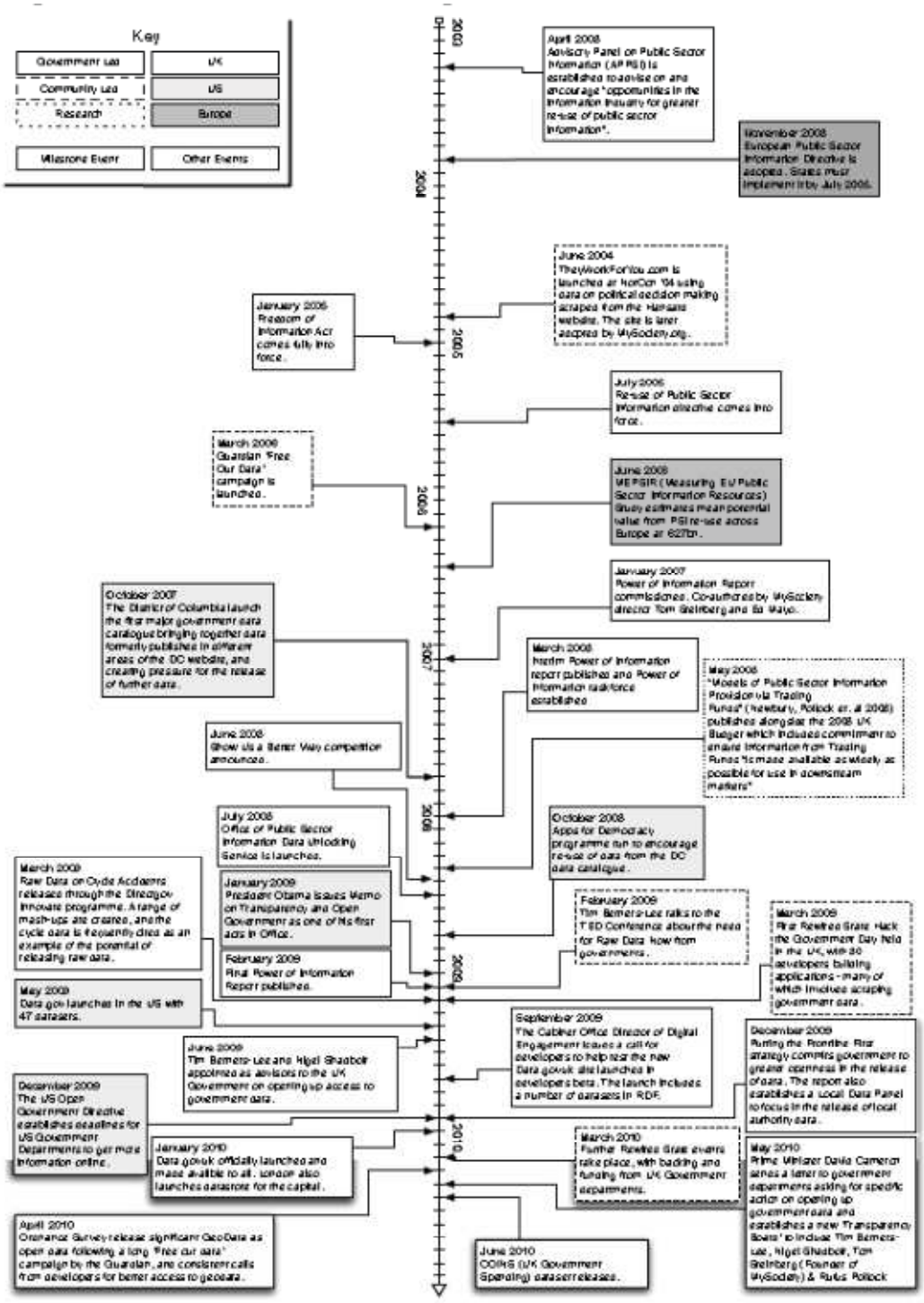
Το συμπέρασμα που βγαίνει από ότι έχει ειπωθεί προηγουμένως είναι ότι ο βασικός λόγος που κάποιος, πολίτης ή επιχείρηση, ασχολείται με τα ανοιχτά κυβερνητικά δεδομένα είναι η επαναχρησιμοποίηση τους, ώστε να προστεθεί αξία στην πρωτογενή μορφή των δεδομένων όπως αυτά εκδίδονται από τις δημόσιες υπηρεσίες. Αυτή η τάση, «κίνημα» για τα ανοιχτά δεδομένα, αυξάνεται τα τελευταία χρόνια διεθνώς και δημιουργικοί εξερευνητές ενός νέου είδους, προσπαθούν να αποκτήσουν πρόσβαση και βγάλουν νόημα από αυτά μέσω της τεχνολογίας. Ξεκινώντας με έργα στις Ηνωμένες Πολιτείες και το Ηνωμένο Βασίλειο, οι κυβερνήσεις σε όλο τον κόσμο εργάζονται τώρα για το άνοιγμα της πληθώρας των δεδομένων τους σε αναγνώσιμη από τους υπολογιστές μορφή. Στην εικόνα 3.3 έχουν συγκεντρωθεί τα κυριότερα γεγονότα, με βάση τον χρόνο, τα οποία σχετίζονται με τα δημόσια κυβερνητικά δεδομένα στις ΗΠΑ, το Ηνωμένο Βασίλειο και την Ευρώπη.

Με τη διάδοση της τεχνολογίας σε όλους τους τομείς της κυβέρνησης, ο πολλαπλασιασμός των δεδομένων συνεχίζεται με ολοένα αυξανόμενη ταχύτητα. Το άνοιγμα αυτών των δεδομένων στο κοινό υπόσχεται να δημιουργήσει αξία για το κοινό η οποία είναι η εξασφάλιση της διαφάνειας και της υπευθυνότητας, η ενθάρρυνση της καινοτομίας και η οικονομική ανάπτυξη, η καλύτερη εκπαίδευση και η βελτίωση της αποτελεσματικότητας των κυβερνήσεων. Οι αξίες αυτές καθιστούν τα δεδομένα του δημόσιου τομέα (Public Sector Information) σε στρατηγικούς πόρους, δυνητικά σημαντικούς για διαφορετικούς οργανισμούς του δημόσιου τομέα, ιδιωτικές επιχειρήσεις, πανεπιστήμια, πολίτες και οργανώσεις πολιτών. Κατά συνέπεια, όπως φαίνεται και στο χρονοδιάγραμμα, η Ευρωπαϊκή Επιτροπή (EC) εξέδωσε την οδηγία για την επαναχρησιμοποίηση του PSI 2003/98/EC, της οποίας ο σκοπός ήταν να θέσει το πλαίσιο των κανόνων στους κυβερνητικούς οργανισμούς και χωρίς να επιβάλλει την διάθεση των δεδομένων, η οποία παρέμεινε στην διακριτική ευχέρεια των μελών. Οι κανόνες αφορούν τις μορφές διαθεσιμότητας με ηλεκτρονικά μέσα, το ανώτατο όριο χρέωσης βασιζόμενο στις δαπάνες για την παραγωγή των δεδομένων, τη διαφάνεια των όρων που ισχύουν για επαναχρησιμοποίηση και τη διαθεσιμότητα των αδειών οι οποίες δεν πρέπει να περιορίζουν τις δυνατότητες για περαιτέρω χρησιμοποίηση των δεδομένων (European Commission 2003).

Η Katleen Jenssen (2011) κάνοντας μια ανασκόπηση στην Ευρωπαϊκή οδηγία, η οποία έπρεπε να υιοθετηθεί με νόμο από τα κράτη μέλη μέχρι το 2005, αναφέρει ότι το 2008 η εφαρμογή και ο αντίκτυπος της οδηγίας επανεξετάστηκε και έδειξε ότι η οδηγία είχε θετική επίπτωση στα κράτη μέλη και στους περισσότερους ενδιαφερόμενους αλλά η πραγματική επίπτωση ήταν ακόμη ασαφής λόγω της πρόσφατης μεταφοράς της οδηγίας στα περισσότερα κράτη-μέλη (European Commission, 2009).

Το νομικό πλαίσιο είναι σημαντικός παράγοντας για την διευκόλυνση της επαναχρησιμοποίησης των δεδομένων του δημόσιου τομέα (PSI). Τα κράτη-μέλη της Ευρωπαϊκής Ένωσης, άλλα αργότερα και άλλα πιο σύντομα, εφάρμοσαν την οδηγία σε νόμο, με την Σουηδία και την Ιταλία να είναι οι τελευταίες χώρες που το έκαναν πράξη το 2010. Οι Schellong και Stepanets (2011) αναφέρουν με ποιό τρόπο τα Ευρωπαϊκά κράτη, μέσω δυσκολιών που συνάντησαν στους νόμους τους και σε

διάφορα άλλα θέματα που αφορούσαν τα ίδια τα δεδομένα, κατάφεραν τελικά να εναρμονιστούν με την Ευρωπαϊκή οδηγία.



Εικόνα 3.3 Χρονοδιάγραμμα επιλεγμένων γεγονότων σχετικά με τα Δημόσια Κυβερνητικά Δεδομένα (Davies T., 2010)

Τέλος η Ευρωπαϊκή Επιτροπή στην «Digital Agenda for Europe» (European Commission, 2010) προσδιόρισε την επαναχρησιμοποίηση των Δημόσιων Κυβερνητικών Δεδομένων, μαζί με την εύκολη και γρήγορη πρόσβαση στο διαδίκτυο, ως το κλειδί για την υλοποίηση μιας ενιαίας ψηφιακής αγοράς και στις 12/12/2011 εξέδωσε ένα δελτίο τύπου (European Commission 2011) το οποίο τιτλοφορείται «Digital Agenda: Turning government data into gold» και αναφέρει ότι η στρατηγική της για τα Δημόσια Ανοιχτά Δεδομένα αναμένεται να αποφέρει μια ώθηση ύψους 40 εκατομμυρίων ευρώ στην Ευρωπαϊκή οικονομία κάθε χρόνο.

3.4.3. ΕΜΠΟΡΙΚΗ ΑΞΙΑ ΤΩΝ ΑΝΟΙΧΤΩΝ ΔΗΜΟΣΙΩΝ ΔΕΔΟΜΕΝΩΝ

Η έρευνα που έγινε για την εμπορική αξία των δεδομένων του δημόσιου τομέα (PIRA, 2000) ήταν η πρώτη Ευρωπαϊκή έρευνα η οποία παρείχε συγκριτικές πληροφορίες για την αξία των αγορών του PSI και την συνεισφορά του PSI οικονομική δραστηριότητα. Η μελέτη εντόπισε δύο κύριες εκτιμήσεις για την αξία των δεδομένων, την αξία των επενδύσεων και την οικονομική αξία και υπολόγισε την συνολική αξία των δεδομένων στα 68 δισεκατομμύρια ευρώ.

Μετά την παραπάνω έρευνα ακολούθησε η MERSIP (Measuring European Public Sector Information Resources) από τους Dekkers et al. (2006), η οποία ανέπτυξε και δοκίμασε μια επαναλαμβανόμενη μεθοδολογία για τη μέτρηση της επαναχρησιμοποίησης των δεδομένων και υιοθέτησε μια βασική μέτρηση της περαιτέρω χρήσης τους στην Ευρωπαϊκή Ένωση (EU 25) και τη Νορβηγία, και σε σύγκριση με τις Ηνωμένες Πολιτείες.

Τα δεδομένα που εξετάστηκαν ήταν από τις παρακάτω έξι κατηγορίες:

- Γεωγραφικά δεδομένα όλων των ειδών
- Μετεωρολογικά δεδομένα
- Επιχειρηματικά δεδομένα, συμπεριλαμβανομένων των διπλωμάτων ευρεσιτεχνίας, των εμπορικών σημάτων και των δημόσιων βάσεων δεδομένων
- Κοινωνικά δεδομένα, συμπεριλαμβανομένων των οικονομικών, της απασχόλησης, της υγείας, του πληθυσμού, της δημόσιας διοίκησης και των κοινωνικών στατιστικών

- Δεδομένα για τη μεταφορά
- Νομικά δεδομένα, συμπεριλαμβανομένων των αποφάσεων εθνικών, αλλοδαπών και διεθνών δικαστηρίων, την εθνική νομοθεσία και τις συνθήκες.

Δεν περιλαμβάνουν επιστημονικά και ερευνητικά δεδομένα ή αυτά με πολιτιστικό περιεχόμενο. Συγκεντρώθηκαν δεδομένα για τη μελέτη από τα μέσα του 2005 μέχρι τις αρχές του 2006 και οι εκτιμήσεις αντιπροσωπεύουν την κατάσταση στις αρχές του 2006. Το αποτέλεσμα του υπολογισμού για την συνολική αγορά των ανοιχτών κυβερνητικών δεδομένων ήταν περίπου 27 εκατομμύρια ευρώ με ανώτατο όριο τα 46 περίπου εκατομμύρια ευρώ.

Την αύξηση της ευημερίας της κοινωνίας και των πολιτών του Ηνωμένου Βασιλείου και τα συνολικά οικονομικά οφέλη από το άνοιγμα των κυβερνητικών δεδομένων προς επαναχρησιμοποίηση υπολογίζει ο Pollock (2011, 2008). Η εκτίμηση για τα κέρδη από την διάθεση των δεδομένων που η τιμή τους κινείται στο οριακό κόστος (ουσιαστικά μηδενικό κόστος) υπολογίζονται στο διάστημα από 4,5 έως 6 δισεκατομμύρια ευρώ ανά έτος. Επισημαίνει επίσης ότι υπάρχει ένα ευρύ φάσμα ωφελειών που θα προκύψουν από το άνοιγμα της πρόσβασης στα δεδομένα. Αυτά περιλαμβάνουν την ανάπτυξη νέων προϊόντων χτισμένα ακριβώς πάνω στα δημόσια δεδομένα, την ανάπτυξη συμπληρωματικών προϊόντων, όπως νέο λογισμικό και υπηρεσίες, τη μείωση του κόστους των συναλλαγών όσον αφορά την πρόσβαση και τη χρήση αυτών των πληροφοριών, κέρδη στο ίδιο το δημόσιο τομέα κ.α.

Η έρευνα του Vickery G. (2011) συγκεντρώνει και παρουσιάζει τις μελέτες για την εμπορική πλευρά των δημόσιων ανοιχτών δεδομένων καθώς και εκτιμήσεις για την αξία της επαναχρησιμοποίησης των δεδομένων στην Ευρώπη. Επίσης κατηγοριοποιεί τα Ανοιχτά Δημόσια Δεδομένα με παραδείγματα όπως φαίνεται στον Πίνακα 3.1. Αυτή η λίστα δεν είναι ούτε δεσμευτική ούτε αποκλειστική. Εξάλλου, είναι δύσκολο να γίνει σαφής διαχωρισμός μεταξύ πολιτιστικού, εκπαιδευτικού και επιστημονικού περιεχομένου. Οι τύποι δεδομένων που χρησιμοποιούνται συνήθως στις εμπορικές εφαρμογές είναι γεωγραφικά, μετεωρολογικά, επιχειρηματικά και οικονομικά, κοινωνικά και μεταφορών καθώς

επίσης και μερικά από το νομικό σύστημα πληροφοριών. Πολιτιστικές, εκπαιδευτικές και επιστημονικές και πολιτικές πληροφορίες είναι συχνά και είναι ευρέως διαθέσιμες από τις κυβερνήσεις. Αλλά, όπως φαίνεται στον πίνακα 3.1 οι διάφοροι τομείς είναι μια συνέχεια των παραδειγμάτων και δεν αλληλοαναιρούνται (Vickery, 2011).

Commercial re-use of PSI	*Geographic Information*	*cartographic information*
		land use info (cadastral data)
		spatial data/geographical coordinates
		administrative and political boundaries
		topographical information
		elevation data
	Meterological and Environmental Information	*oceanographic data*
		hydrographic data
		environmental (quality) data
		atmospheric data
		meteorological (weather) data
	Economic and Business Information	*financial information*
		company information
		economic and statistics
		industry and trade information
	Social Information	*demographic information*
		attitude surveys
		data on health/illness
		education and labour statistics
	Traffic and Transport Information	*transport network information*
		traffic information
		transport statistics
		car registration data
	Tourist and Leisure Information	*hotel information*
		tourism statistics
		entertainment (local and national)
	Agricultural, Farming, Forestry and Fisheries information	*cropping/land use data*
		farm incomes/use of resources
		fish farming/harvest information
		live stock data
	Natural Resource Information	*biologic and ecologic information*
		energy resource/consumption information
		geological and geophysical information
	Legal System Information	*crime/conviction data*
		laws
		information on rights and duties
		information on legislation
		information on judicial decisions
		patent and trademark information
	Scientific Information and Research data	*university research*
		publicly-funded research institutes
		governmental research
	Educational Content	*academic papers and studies*
		lecture material
	Political Content	*governmental press releases*
		local and national proceedings of governments
		green papers
Making available PSC	*Cultural Content*	*museum material*
		gallery material
		archeological sites
		library resources
		public service broadcast archives
		other public archives

Πίνακας 3.1 Κατηγορίες Ανοιχτών Δημόσιων Δεδομένων με παραδείγματα. (Vickery, 2011)

Συμπερασματικά μπορούμε να σημειώσουμε, βλέποντας τις εκτιμήσεις των μελετών, ότι η επαναχρησιμοποίηση των ανοιχτών κυβερνητικών δεδομένων έχει

την δυναμική να ωθήσει οικονομικά τις οικονομίες των χωρών της Ευρωπαϊκής Ένωσης, αλλά και ολόκληρης της υφηλίου. Καινοτόμες εφαρμογές και τεχνολογίες μπορούν να κατασκευαστούν «πάνω» στα δεδομένα οι οποίες θα εμπλέξουν τους πολίτες σε συνεργασία με τις κυβερνήσεις και θα φωτίσουν και θα βελτιώσουν τις εσωτερικές δημόσιες λειτουργίες. Η ενασχόληση του ιδιωτικού τομέα στην μεταφορά των δεδομένων προς τους πολίτες θα μπορούσε να υιοθετηθεί ως κυβερνητική στρατηγική καθώς ο ιδιωτικός τομέας είναι πιο ευέλικτος και ικανός να παρουσιάσει τελικά τα κυβερνητικά δεδομένα και αυτή η στρατηγική έχει ως αποτέλεσμα την μείωση του συνολικού κόστους τεχνολογικής ανάπτυξης από την κυβέρνηση (Robinson 2009; Burdon 2009; Pizzicannella, 2010).

3.4.4. ΜΟΡΦΕΣ ΕΠΑΝΑΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

Η επαναχρησιμοποίηση των κυβερνητικών δεδομένων είναι στην ουσία ο άξονας και ο τρόπος με τον οποίο αυτά παίρνουν αξία. Η μετατροπή των πρωτογενών κυβερνητικών στοιχείων σε δομημένη πληροφορία είναι αυτή που δίνει την εικόνα, με διάφορους τρόπους, των στοιχείων που συλλέγουν οι δημόσιες υπηρεσίες. Όμως για να μπορέσει κάποιος να χρησιμοποιήσει τα δεδομένα, μετατρέποντας τα σε πληροφορία ή χτίζοντας εφαρμογές που στηρίζονται σ` αυτά, πρέπει αυτά να είναι διαθέσιμα ηλεκτρονικά και στην κατάλληλη μορφή.

Ανατρέχοντας στις αρχές των Ανοιχτών Κυβερνητικών Δεδομένων παρατηρούμε ότι αυτά για να θεωρηθούν ανοιχτά πρέπει να μπορούν να επεξεργασθούν από υπολογιστή και να είναι σε μη αποκλειστική μορφή (ανοιχτά πρότυπα). Ο Cyganiac (2011) κάνει μια αρχική κατηγοριοποίηση των μορφών των δεδομένων στις παρακάτω κατηγορίες η οποία είναι αρκετά απλοϊκή, όμως μας δίνει μία πρώτη εικόνα για το ποιες μορφές είναι φιλικές προς περαιτέρω χρησιμοποίηση:

- **Καλή:** MS Excel, CSV, XML, JSON, Microdata
- **Όχι τόσο καλή:** Websites, MS Word
- **Κακή:** PDF
- **Πολύ κακή:** Μόνο διαγράμματα, χάρτες χωρίς νούμερα

Το πόσο καλή ή κακή είναι η μορφή των δεδομένων, σύμφωνα με την παραπάνω κατηγοριοποίηση, αφορά την αυτοματοποιημένη επεξεργασιμότητα των δεδομένων από υπολογιστή.

Στον πίνακα 3.2 εμφανίζονται οι περισσότερες μορφές αρχείων χωρισμένες σε αρχεία κειμένου, πίνακες και φωτογραφίες από την μία πλευρά και αρχεία γεωχωρικών δεδομένων από την άλλη και αυτά χαρακτηρίζονται από την επεξεργασιμότητα, τις προδιαγραφές και αν είναι κάτω από ανοιχτά πρότυπα. Η έννοια «επεξεργασιμότητα από υπολογιστές» εξηγείται από την Βρετανική COI (Central Office of Information, 2011) ως εξής: «Στο πλαίσιο της δημοσίευσης των δεδομένων, επεξεργασιμότητα από υπολογιστή σημαίνει να γίνονται τα βασικά δεδομένα, τα οποία δημοσιεύονται, προσβάσιμα για χρήση μέσω διαδικασιών βασισμένων σε υπολογιστή χωρίς να απαιτείται η ανθρώπινη ερμηνεία.»

Επίσης τονίζεται ότι τα ανοιχτά κυβερνητικά δεδομένα πρέπει να είναι κάτω από ανοιχτά πρότυπα γιατί αυτά δεν βασίζονται σε ένα μόνο προϊόν ή επιχείρηση, είναι ελεύθερα διαθέσιμα στο κοινό χωρίς καμία άδεια, δίπλωμα ευρεσιτεχνίας ή άλλους περιορισμούς και μεγιστοποιούν τις πιθανές χρήσεις, όπως σημειώνεται και στις αρχές που είδαμε σε προηγούμενο κεφάλαιο.

Οι προδιαγραφές διευκρινίζονται από τον οδηγό του Cornell University Library`s Digital Preservation Management ως αυτές που υποδεικνύουν την σωστή υποκατηγορία, κωδικοποίηση, ακολουθία, διάταξη, μέγεθος και εσωτερικές σχέσεις που προσδιορίζουν μοναδικά την συγκεκριμένη μορφή του αρχείου και επιτρέπουν την κατανόηση του και την χρήση του. Για παράδειγμα, μια προδιαγραφή για τη μορφή θα πρέπει να αναφέρει τη θέση των ουσιαστικών ορίων εντός των bitstream και αν μια συγκεκριμένη υπομονάδα πρέπει να ερμηνευθεί ως ASCII χαρακτήρας, αριθμητική τιμή, εντολή μηχανής, επιλογή χρωμάτων ή κάτι άλλο (Cornell University , [n.d.]).

Στον πίνακα 3.3 (Παπαδόπουλος, 2011) ο συγγραφέας προσθέτει και μία άλλη διάσταση των ανοιχτών δεδομένων την οποία ονομάζει «ελεύθερο», αναφερόμενος στα πρότυπα διάθεσης των δεδομένων. Όταν λέμε ότι το πρότυπο είναι ελεύθερο εννοούμε ότι είναι ανοιχτό και επιπλέον έχει αποδοθεί στο public domain έτσι ώστε να μην υπάρχει κανενός είδους πατέντας ή παρακρατημένο

δικαίωμα και να μην υφίσταται δυνατότητα ανάκλησης ή τροποποίησης των όρων διάθεσής του.

Text, Tables and Pictures	M	S	O	Geodata	M	S	O
Text (.txt)	+	+	+	Geography Markup Language (.gml)	+	+	+
Comma Separated Value (.csv)	+	+	+	GPS Exchange Format (.gpx)	+	+	+
Hypertext Markup Language (.html)	+	+	+	Keyhole Markup Language (.kml)	+	+	+
Extensible Markup Lanuage (.xml)	+	+	+	Drawing Interchange File Format (.dxf)	+	+	-
Resource Description Framework (.rdf)	+	+	+	Autodesk Drawing Format (.dwg)	+	-	-
Open Document Formats (.odt, .ods..)	+	+	+	ESRI Shapefile Format (.shp, .shx, .dbf)	+	+	-
Newsfeed / Webfeed Syndication (.rss)	+	+	+	Enhanced Compression Wavelet (.ecw)	+	-	-
Portable Document Format (.pdf)	-	+	+	MrSID Format (.sid)	+	-	-
Microsoft Word (.doc, .docx)	-	+	-	Normbas. Exchangeinterface (NAS)	+	+	-
Microsoft Excel (.xls, .xlsx)	+	+	-	Unified Database Interface (EDBS)	+	+	-
Microsoft Rich Text Format (.rtf)	+	+	-	BGRUND (Geospital agency BaWü)	+	+	-
Graphics Interchange Format (.gif)	-	+	-	WLDGE-Format (.wld)	+	+	-
JPEG Format (.jpg / .jpeg)	-	+	+				
Portable Networks Graphics (.png)	-	+	+				
Tagged Image File Format (.tiff, .tif)	-	+	-				
Geo-TIFF Format (.geotiff)	-	+	-				

M = machine-readable; S = specifications available; O = open format[26]

Πίνακας 3.2 Πιθανές μορφές δεδομένων για Ανοιχτά Κυβερνητικά Δεδομένα (Schellong, 2011)

	Επαναχρησιμοποιήσιμο			Ανοιχτό	Ελεύθερο
	Μηχανικά αναγνώσιμο	Μηχανικά κατανοητό (για δεδομένα)	Διαθέσιμες προδιαγραφές	Άδεια χωρίς περιορισμούς	
Plain Text (.txt)	√	×	√	√	√
Comma Seperated Value (.csv)	√	√	√	√	√
Hyper Text markup (.html / .html)	√	×	√	√	√
Extensible Markup Language (.xml)	√	√	√	√	√
Resource Description Framework (.rdf)	√	√	√	√	√
Open Document Format (.odt)	√	×	√	√	×
Open Document Format (.ods)	√	√	√	√	×
Microsoft Office format (.doc)	√	×	√	√	×
Microsoft Office format (.xls)	√	√	√	√	×
Microsoft Office 2007+ (.docx)	√	×	√	√	×
Microsoft Office 2007+ (.xlsx)	√	√	√	√	×
Portable Document Format (.pdf)	×	×	√	√	×
Image File (.jpg / .png)	×	×	√	√	-

Πίνακας 3.3 Πρότυπα διάθεσης δεδομένων (Παπαδόπουλος, 2011)

Τέτοια πρότυπα αποτελούν κάποια στάνταρ βιομηχανικά πρότυπα (όπως το .csv) καθώς και πρότυπα που έχουν δημιουργηθεί από διεθνείς οργανισμούς προτυποποίησης (όπως το .xml).

3.5. ΤΡΟΠΟΙ ΔΗΜΟΣΙΕΥΣΗΣ ΤΩΝ ΑΝΟΙΧΤΩΝ ΚΥΒΕΡΝΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

3.5.1. ΕΙΣΑΓΩΓΗ

Οι κυβερνητικές πρωτοβουλίες δημοσιοποίησης των ανοιχτών δεδομένων πληθαίνουν το τελευταίο χρονικό διάστημα με συνέπεια οι κατάλογοι στους οποίους φιλοξενούνται κυβερνητικά δεδομένα να αυξάνονται και παράλληλα να αυξάνονται και οι επιλογές αυτών που θέλουν να χρησιμοποιήσουν τα δεδομένα. Υπάρχουν διαφορετικοί τρόποι ώστε τα δεδομένα να φτάσουν τελικά στους πολίτες και ένας από αυτούς είναι τα συνδεδεμένα δεδομένα που πρότεινε ο εφευρέτης του Παγκόσμιου Ιστού των Εγγράφων Sir Tim Berners-Lee. Μπορούμε επίσης να βρούμε εφαρμογές που ποικίλλουν ανάλογα με το είδος των συνόλων δεδομένων που χρησιμοποιούν αλλά και το πνεύμα της καινοτομίας των δημιουργών τους.

3.5.2. ΠΡΑΚΤΙΚΕΣ ΔΗΜΟΣΙΟΠΟΙΗΣΗΣ

Η μορφή δημοσιοποίησης των δεδομένων, σύμφωνα με τα παραπάνω, παίζει σημαντικό ρόλο στην δυνατότητα επαναχρησιμοποίησης τους. Οι πίνακες που είδαμε στο προηγούμενο κομμάτι εκτός από το να κατατάσσουν τις μορφές των δεδομένων σύμφωνα με το ποιές είναι πιο φιλικές για περαιτέρω χρήση από οποιονδήποτε θα ήθελε να το κάνει, μας δείχνουν και ότι τα δεδομένα μπορούν να δημοσιοποιηθούν σε διάφορες μορφές οι οποίες να είναι «καλές» για να δημιουργήσει κάποιος χρήσιμες πληροφορίες από αυτά. Έτσι σύμφωνα με την Βρετανική COI (Central Office of Information, 2011) τα .csv αρχεία είναι κατάλληλα για την δημοσιοποίηση πινάκων βάσεων δεδομένων, τα .rdfa αρχεία χρησιμοποιούνται για ιστοσελίδες που επιβάλλεται να έχουν μεταδεδομένα και η ίδια ιστοσελίδα δεδομένων μπορεί να διαβαστεί και από ανθρώπους και από μηχανές και τα .xml αρχεία είναι κατάλληλα για πίνακες και μαθηματικά.

Οι Bennett και Harvey (2009) προτείνουν μία σειρά από βήματα τα οποία οι κυβερνητικοί οργανισμοί θα πρέπει να έχουν ως προτεραιότητα για την επίτευξη καλής πρακτικής δημοσιοποίησης ανοιχτών δεδομένων:

- Δημοσιοποίηση των δεδομένων στην ακατέργαστη μορφή τους, αλλά τα δεδομένα πρέπει να είναι σωστά δομημένα για να μπορεί κάποιος επιτυχώς να τα χρησιμοποιήσει αυτοματοποιημένα. Τα σωστά δομημένα δεδομένα αναφέρονται στην μορφή τους και περιλαμβάνουν μορφές που συζητήσαμε προηγουμένως όπως .xml, .rdf και .csv αρχεία.
- Δημιουργία online καταλόγων των δεδομένων, μαζί με τεκμηρίωση, έτσι ώστε κάποιος να μπορεί να βρει τα δεδομένα που έχουν δημοσιοποιηθεί.
- Δημιουργία των δεδομένων σε μορφή κατανοητή και από ανθρώπους και από υπολογιστές εμπλουτίζοντας τους υπάρχοντες (X)HTML πόρους με μεταδεδομένα και αναγνωριστικά, κωδικοποιώντας τα δεδομένα με την χρήση ανοιχτών βιομηχανικών προτύπων, χρήση μόνιμων μοντέλων και ανιχνεύσιμων «cool URIs» (http://www.w3.org/TR/cooluris/) και με την ύπαρξη ηλεκτρονικών αναφορών με την μορφή υπερσυνδέσμων.

Στον ίδιο άξονα κινείται και ο Tauberer (2009) προσθέτοντας στα παραπάνω και την δημιουργία εφαρμογών και web υπηρεσιών οι οποίες θα επιτρέπουν σε κάποιον να αναζητεί και να ανακτά αυτοματοποιημένα αυτό που ψάχνει καθώς και να υποβάλλει πληροφορίες.

3.5.3. ΣΥΝΔΕΔΕΜΕΝΑ (ΚΥΒΕΡΝΗΤΙΚΑ) ΑΝΟΙΧΤΑ ΔΕΔΟΜΕΝΑ

Στο προηγούμενο κομμάτι μιλήσαμε για πρακτικές δημοσιοποίησης, αναφέροντας την χρήση μοντέλων και αναγνωριστικών (URIs). Η πρακτική αυτή δημοσίευσης δεδομένων που χρησιμοποιεί τα URIs (Unified Resource Identifiers) ονομάζεται συνδεδεμένα δεδομένα (linked data) και χρησιμοποιείται για να συνδέει δομημένα σύνολα δεδομένων στο διαδίκτυο. Το διαδίκτυο άλλαξε ριζικά τον τρόπο με τον οποίο μοιράζεται η γνώση, βοηθώντας την δημοσίευση εγγράφων και την πρόσβαση σ` αυτά ως ένα κομμάτι ενός παγκόσμιου πληροφοριακού συστήματος. Παρά τα αδιαμφισβήτητα οφέλη που παρέχει το διαδίκτυο, μέχρι πρόσφατα οι

αρχές που υποβοήθησαν τον παγκόσμιο ιστό των εγγράφων να αναπτυχθεί δεν βρήκαν εφαρμογή στα δεδομένα. Τα τελευταία χρόνια όμως το διαδίκτυο έχει εξελιχθεί από ένα παγκόσμιο χώρο από συνδεδεμένα έγγραφα σε έναν χώρο όπου και τα έγγραφα και τα δεδομένα είναι συνδεδεμένα. Η στήριξη αυτής της εξέλιξης είναι η υιοθέτηση ενός συνόλου βέλτιστων πρακτικών για την δημοσίευση και την σύνδεση δομημένων δεδομένων στον Παγκόσμιο Ιστό που οδηγεί στα συνδεδεμένα δεδομένα και το σημασιολογικό Ιστό (Semantic Web) (Bitzer et al., 2009).

Ο όρος «linked data» αναφέρθηκε από τον Sir Tim Berners-Lee (2006) σ` ένα σχεδιαστικό σημείωμα, όπου ουσιαστικά πρότεινε την εφαρμογή των ίδιων κανόνων που έκαναν επιτυχημένο τον Ιστό των Εγγράφων στα δεδομένα, καθώς ο ίδιος είναι ο εφευρέτης του Παγκόσμιου Ιστού (World Wide Web). Την περίοδο εκείνη λοιπόν οι κανόνες εκφράστηκαν με τον παρακάτω τρόπο:

- Όλα τα αντικείμενα πρέπει να ταυτοποιούνται με τη χρήση URIs για ονόματα.
- Να χρησιμοποιούνται HTTP URIs ώστε κάποιος να μπορεί να ψάξει αυτά τα ονόματα.
- Όταν κάποιος ανατρέχει σε ένα URI, δηλαδή σε μια RDF ιδιότητα που μεταφράζεται ως υπερσύνδεσμος, οδηγείται σε περισσότερα χρήσιμα δεδομένα.
- Πρέπει να συμπεριλαμβάνονται σύνδεσμοι προς άλλα URIs ώστε να επιτρέπουν την ανακάλυψη περισσότερων δεδομένων.

Ο όρος συνδεδεμένα δεδομένα σύμφωνα με τον Bitzer et al. (2009) αναφέρεται σε «*δεδομένα δημοσιευμένα στο διαδίκτυο με τρόπο ώστε να είναι αναγνώσιμα από υπολογιστές, το νόημα τους να είναι ρητά ορισμένο, να είναι συνδεδεμένα με άλλα εξωτερικά σύνολα δεδομένων και να μπορούν να συνδεθούν από εξωτερικά σύνολα δεδομένων*».

Το πιο σημαντικό παράδειγμα υιοθέτησης και εφαρμογής των αρχών των συνδεδεμένων δεδομένων είναι το Linking Open Data Project, μία προσπάθεια που ξεκίνησε τον Ιανουάριο του 2007. Ο αρχικός και συνεχιζόμενος σκοπός του project είναι να ξεκινήσει τον ιστό των δεδομένων (web of data), προσδιορίζοντας τα υπάρχοντα σύνολα δεδομένων που έχουν άδεια, μετατρέποντάς τα σε rdf, σύμφωνα

με τις αρχές των συνδεδεμένων δεδομένων και δημοσιεύοντάς τα στο διαδίκτυο. Μέχρι το 2008, ερευνητές και αυτοί που πρώτοι υιοθέτησαν τις πρακτικές των συνδεδεμένων δεδομένων αντάλλασαν συμβουλές σχετικά με τη δημοσίευση μεγάλων συνόλων δεδομένων ως Linked Data, αναζητώντας δομημένο περιεχόμενο στον Παγκόσμιο Ιστό, και αυτό που χρειαζόταν ως συνδεδεμένες εφαρμογές δεδομένων. Το LOD project συνέχισε να μεγαλώνει με την συμμετοχή μεγάλων οργανισμών, όπως το BBC, το Thomson Reuters και η βιβλιοθήκη του Κογκρέσου. Μία εικόνα του μεγέθους και της αύξησή του βλέπουμε στις εικόνες 3.4 και 3.5.

Από τότε μέχρι το 2011 πάνω από 300 σύνολα δεδομένων έχουν δημοσιευθεί στο Διαδίκτυο, ως συνδεδεμένα δεδομένα, αξιολογήσεις και βιβλία γενικά έχουν γραφτεί με θέμα τα Linked Data, και ξεκίνησε μια ομάδα εργασίας του W3C εστιάζοντας σε Δημόσια Κυβερνητικά Δεδομένα. Σε αυτό το σύντομο χρονικό διάστημα τα συνδεδεμένα δεδομένα έχουν εξελιχθεί σε μια κύρια δραστηριότητα για πολλές κυβερνήσεις σε όλο τον κόσμο.

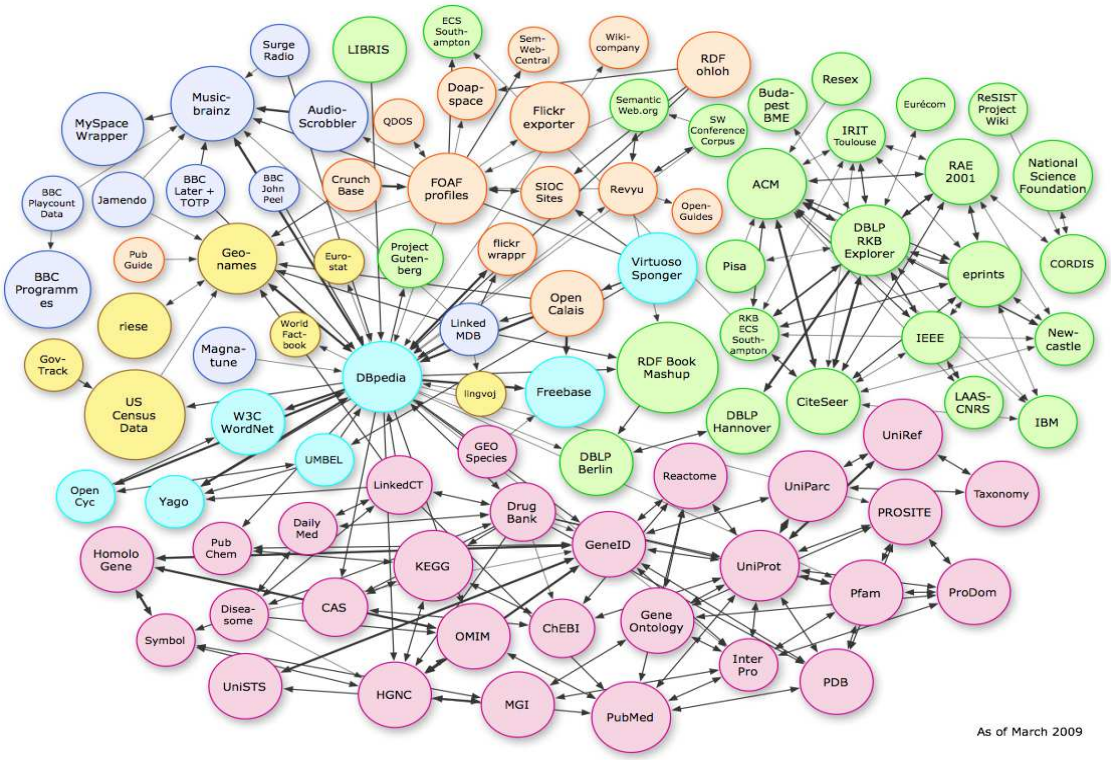
Στην ιστοσελίδα των δεδομένων της κυβέρνησης των ΗΠΑ περιλαμβάνονται είκοσι μία χώρες των οποίων οι κυβερνήσεις έχουν δημοσιεύσει δεδομένα σχετικά με τη λειτουργία του δημόσιου τομέα τους, με το ένα τρίτο του από αυτά να καταχωρούνται ως συνδεδεμένα δεδομένα χρησιμοποιώντας τα πρότυπα του World Wide Web Consortium. Αυτοί οι αριθμοί δείχνουν σημάδια αύξησης στο άμεσο μέλλον (Hyland et al., 2011).

Ο Sir Tim Berners-Lee (2010) πρότεινε ένα μοντέλο κατάταξης για να ενθαρρύνει τις δημόσιες υπηρεσίες ώστε να δημοσιεύσουν τα δεδομένα τους. Το μοντέλο αυτό των πέντε αστεριών είναι:

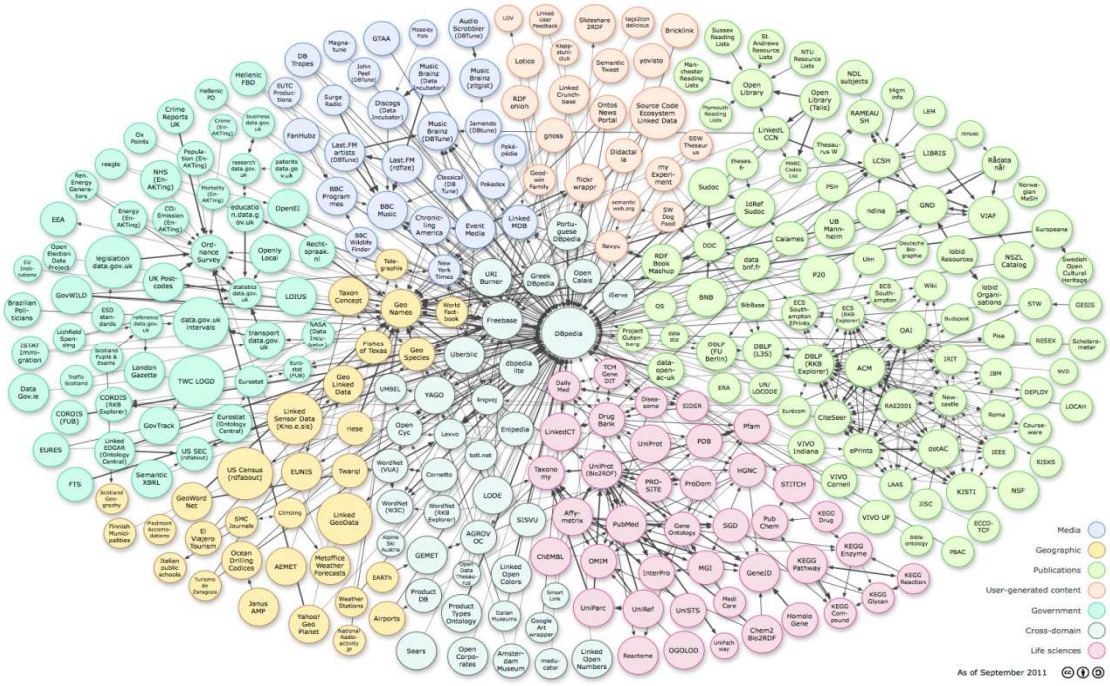
★ Δημοσίευση των δεδομένων στο διαδίκτυο σε οποιαδήποτε μορφή, αλλά κάτω από ανοιχτή άδεια.

★★ Δημοσίευση των δεδομένων σε μορφή που είναι δομημένη και αναγνώσιμη από υπολογιστές (π.χ. excel αντί μιας σκαναρισμένης εικόνας).

★★★ Προσθέτουμε στα παραπάνω την δημοσίευση των δεδομένων σε μη ιδιωτικές μορφές (π.χ. .csv αντί .xls).



Εικόνα 3.4 Διάγραμμα συνδεδεμένων συνόλων δεδομένων τον Μάρτιο 2009 (Bitzer, et al., 2009)



Εικόνα 3.5 Διάγραμμα συνδεδεμένων συνόλων δεδομένων τον Σεπτέμβριο 2011. (http://richard.cyganiak.de/2007/10/lod/)

★★★★ Προσθέτουμε στα παραπάνω την δημοσίευση των δεδομένων με ανοιχτά πρότυπα από το W3C (.rdf).

★★★★★Σύνδεση των διαθέσιμων δεδομένων.

Σύμφωνα λοιπόν με τις αρχές των συνδεδεμένων δεδομένων και με τα τρία βήματα που προτείνει το W3C για τις δημόσιες υπηρεσίες να δημοσιοποιούν τα δεδομένα τους, η Βρετανική κυβέρνηση δεσμεύτηκε να δημοσιεύσει τα περισσότερα από τα δεδομένα της ως συνδεδεμένα μέχρι το τέλος του 2011 (HM Government, 2009).

3.5.4. ΚΑΤΑΛΟΓΟΙ ΔΗΜΟΣΙΩΝ ΚΥΒΕΡΝΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Ένα από τα τρία βήματα που προτείνουν οι Bennett και Harvey (2009) είναι η δημιουργία online καταλόγων των δεδομένων, μαζί με τεκμηρίωση, έτσι ώστε κάποιος να μπορεί να βρει τα δεδομένα που έχουν δημοσιοποιηθεί. Οι κατάλογοι αυτοί είναι προϊόν των πρωτοβουλιών ανοιχτών κυβερνητικών δεδομένων και προσφέρουν τα διαθέσιμα δεδομένα συγκεντρωμένα σε ένα σημείο (ιστοσελίδα) (Veljković, 2011). Οι κατάλογοι με εθνικό, περιφερειακό ή τοπικό πεδίο εφαρμογής λειτουργούν ως one-stop data portals παρέχοντας τα διαθέσιμα σύνολα κυβερνητικών δεδομένων μαζί με τα απαραίτητα μεταδεδομένα (Maali et, al., 2010).

Το φαινόμενο των καταλόγων δεδομένων του δημόσιου τομέα δεν είναι καινούργιο, καθώς στις ΗΠΑ και στην Μ. Βρετανία υπάρχει με την λογική της πρόσβασης από ένα σημείο από την δεκαετία του 1990 (Koumenides et al., 2010).
Οι τρόποι διαχείρισης των συνόλων δεδομένων όταν αναπτύσσονται τέτοιου είδους κατάλογοι είναι δύο, τα αποθετήρια δεδομένων και οι κατάλογοι όπου συγκεντρώνονται σύνδεσμοι που «δείχνουν» προς τα σύνολα δεδομένων (Veljković, 2011; Παπαδόπουλος, 2011). Στην πρώτη περίπτωση έχουμε μία μεγάλη βάση συνόλων δεδομένων, όπου συγκεντρώνονται όλα τα δεδομένα από όλες τις υπηρεσίες και προσφέρονται από ένα portal το οποίο λειτουργεί ως αποθετήριο. Στην δεύτερη περίπτωση τα δεδομένα υπάρχουν στις ιστοσελίδες των δημοσίων υπηρεσιών που τα έχουν παράγει αρχικά, και στον κατάλογο συγκεντρώνονται σύνδεσμοι προς τα δεδομένα.

Οι Kalampokis et al., 2011a χρησιμοποιούν τους όρους άμεση παροχή δεδομένων για τα αποθετήρια και έμμεση παροχή δεδομένων για την περίπτωση συγκέντρωσης των συνδέσμων σε ένα σημείο. Στον σχεδιασμό των πρωτοβουλιών των Ανοιχτών Κυβερνητικών Δεδομένων, όσον αφορά την διαχείριση των δεδομένων, τα κύρια θέματα που πρέπει να ληφθούν υπ` όψιν για να πάρει κάποιος την απόφαση ποιά από τις δύο περιπτώσεις θα ακολουθήσει είναι τα εξής:

- Σε ποιόν ανήκουν τα δεδομένα (ποιός τα διατηρεί); Αυτός μπορεί να είναι είτε η δημόσια υπηρεσία είτε το portal.
- Ποιός δημοσιεύει τα δεδομένα (και πιθανόν τα σχετικά μεταδεδομένα); Αυτός μπορεί να είναι είτε η δημόσια υπηρεσία είτε το portal είτε και τα δύο.

Σύμφωνα με την έρευνα και στις δύο περιπτώσεις πρέπει να υπάρχει προσοχή από τους εμπλεκόμενους, που είναι παραπάνω από ένας, ώστε να μην γίνουν παρωχημένα τα δεδομένα και να υπάρχει συγχρονισμός στην περίπτωση που δημοσιοποιούν δεδομένα και τα δύο μέρη. Στην συνέχεια γίνεται μία ανασκόπηση των υπαρχόντων πρωτοβουλιών οι οποίες κατατάσσονται (πίνακας 3.5) με βάση ένα σχήμα (πίνακας 3.4) το οποίο έχει δύο διαστάσεις την τεχνολογική προσέγγιση και την οργανωσιακή προσέγγιση. Η τεχνολογική διάσταση αναφέρεται στην μορφή των συνόλων δεδομένων και αν αυτά ακολουθούν τις αρχές των συνδεδεμένων δεδομένων και η οργανωσιακή διάσταση στην έμμεση ή άμεση παροχή των δεδομένων αυτών από τις πρωτοβουλίες (Kalampokis et al., 2011a).

Η περαιτέρω κατηγοριοποίηση της τεχνολογικής διάστασης έχει να κάνει πρώτα με το αν τα δεδομένα είναι διαθέσιμα για λήψη από το διαδίκτυο και σε ποιές μορφές διατίθενται, αν είναι δηλαδή αναγνώσιμες από υπολογιστή (π.χ. excel) ή κάτω από ανοιχτά πρότυπα (π.χ. csv). Η δεύτερη κατηγορία εξετάζει αν τα δεδομένα ακολουθούν τους κανόνες των συνδεδεμένων δεδομένων και είναι διαθέσιμα μέσω RESTful APIs και μέσω διεπαφές αναζήτησης SPARQL καθώς και αν συνδέουν μεταξύ τους δεδομένα από διαφορετικά σύνολα δεδομένων.

Στους χάρτες που ακολουθούν (εικόνα 3.6) μπορούμε να διακρίνουμε τις πρωτοβουλίες των Δημόσιων Κυβερνητικών Δεδομένων σε παγκόσμιο επίπεδο

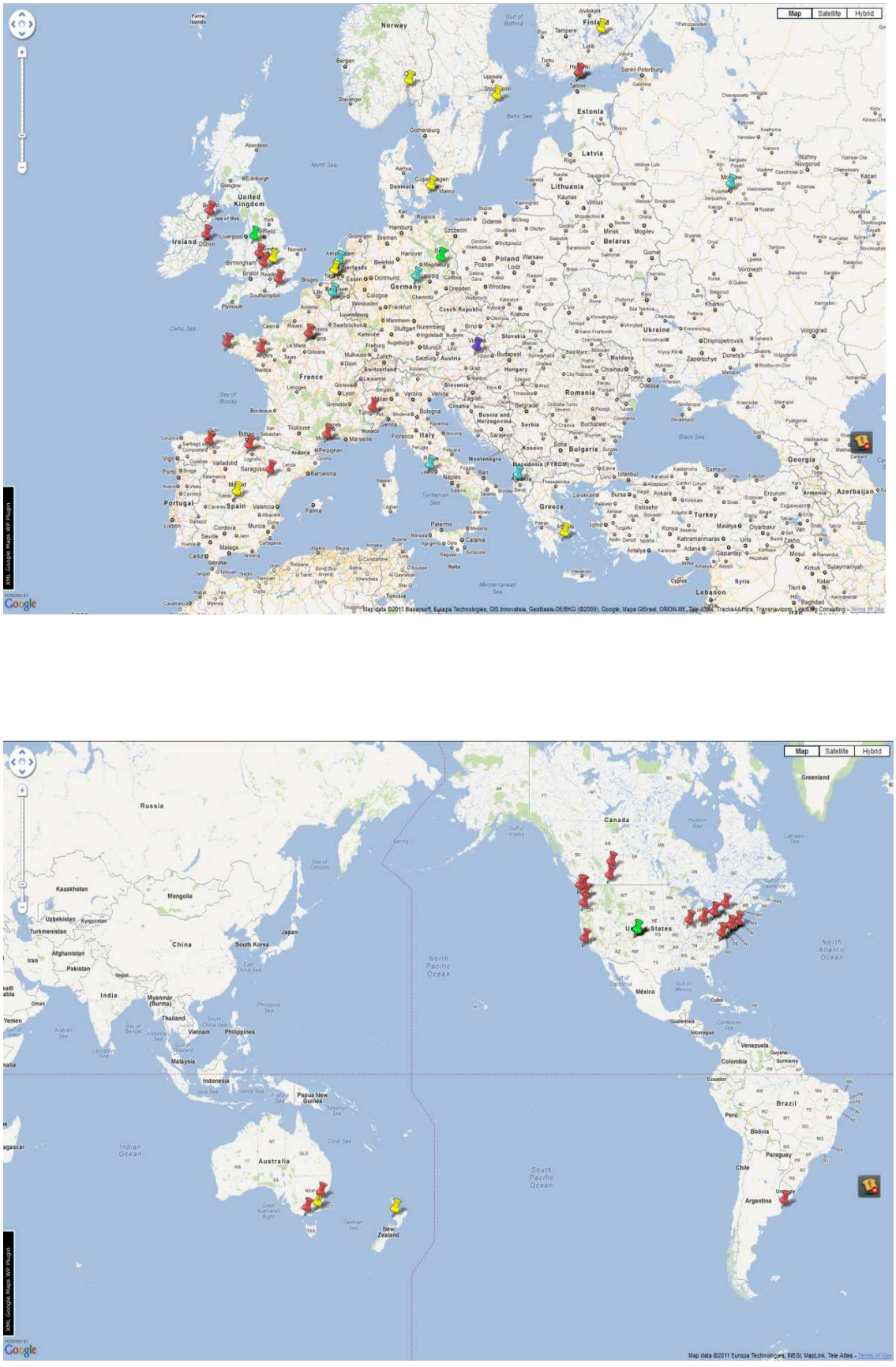
όπως αυτές φαίνονται στην ιστοσελίδα του Open Knowledge Foundation για τα Δ.Κ.Δ (Open Government Data on a Map, 2011).

Technological Approach \ Organizational Approach		Direct Data Provision	Indirect Data Provision
Downloadable Files	Proprietary and desktop-centric formats	Repository of Downloadable Files	Registry of Downloadable Files
	Machine-readable formats		
	Machine-readable formats using open standards		
Linked Data	Linked data principles	Direct Provision of Linked Data	Indirect Provision of Linked Data
	Linking available data		

Πίνακας 3.4 Σχήμα κατάταξης πρωτοβουλιών Δ.Κ.Δ. (Kalampokis et al., 2011a)

Technological Approach \ Organizational Approach		Direct Data Provision	Indirect Data Provision
Downloadable Files	Proprietary and desktop-centric formats		Data.govt.nz Data.nsw Mass.gov/data Open Data Euskadi
	Machine-readable formats	NUC data mine Pic and Mix Toronto.ca/opne Vancouver's open data catalogu	Data.australia.gov.au Data.ca.gov Data.gov.uk Data.govt.nz Data.nsw Data.vic.gov.au DataSF OpendataNI
	Machine-readable formats using open standards	City of Edmonton open data catalogue Data.gov Data.seattle.gov Dati.piemonte.it Dc.gov - data catalogue Lichfield open data London datastore Maine.gov DataShare Toronto.ca/open Vancouver's open data catalogue Warwickshire open data	Data.ca.gov Data.gov.uk Data.vic.gov.au
Linked Data	Linked data principles	Catalogo de Datos de Asturias Data.gov Data.gov.uk	
	Linking available data		Data.gov.uk

Πίνακας 3.5 Κατάταξη υπαρχόντων πρωτοβουλιών σύμφωνα με το σχήμα (Kalampokis et al., 2011a)



Εικόνα 3.6 Παγκόσμιος χάρτης πρωτοβουλιών Δημόσιων Κυβερνητικών Δεδομένων (http://opengovernmentdata.org/data/map/)

Ο παρακάτω πίνακας 3.6 παρουσιάζει σε παγκόσμιο επίπεδο τις πρωτοβουλίες των Δ.Κ.Δ περιλαμβάνοντας πληροφορίες, εκτός από το URL, για την γεωγραφική κάλυψη και την ημερομηνία που «ανέβηκε» στο διαδίκτυο. Στόχος των πρωτοβουλιών αυτών είναι κάτω από εθνική ή τοπική αιγίδα, να προσφέρουνε στον πολίτη και στις επιχειρήσεις τα δεδομένα που συγκεντρώνουν οι δημόσιοι φορείς έτσι ώστε να ενισχυθεί η διαφάνεια, η υπευθυνότητα και η καινοτομία. Οι παραπάνω πίνακες μας δίνουν μία ιδέα για το συμβαίνει σήμερα παγκόσμια στον τομέα των πρωτοβουλιών για το «άνοιγμα» των κυβερνητικών δεδομένων με την δημιουργία online καταλόγων όπου συγκεντρώνονται τα δεδομένα με σκοπό την ευρετηρίαση τους έτσι ώστε να είναι ευκολότερο για κάποιον που ψάχνει τέτοιου είδους δεδομένα, τελικά να μπορεί να τα βρει για να τα επαναχρησιμοποιήσει. Ένα μοντέλο τεσσάρων σταδίων για την μέτρηση της προόδου των πρωτοβουλιών Ανοιχτών Κυβερνητικών Δεδομένων, προτείνεται ως χάρτης από τους Καλαμπόκη et al., 2011b. Ο τρόπος και οι τεχνολογίες που χρησιμοποιούνται για την δημοσιοποίηση διαφέρουν και τελικά η μόνη πρωτοβουλία που έχει δημιουργήσει συνδέσμους μεταξύ διαφορετικών συνόλων δεδομένων είναι αυτή της Βρετανικής Κυβέρνησης http://data.gov.uk.

Εκτός από τις κυβερνητικές πρωτοβουλίες, που είδαμε στον πίνακα 3.4, υπάρχουν και κάποιες άλλες προσπάθειες όπως αυτή του Tetherless World Constellation (TWC) από το Rensselaer Polytechnic Institute (RPI), που ανέπτυξε ένα portal http://logd.tw.rpi.edu/home ακολουθώντας τις αρχές των συνδεδεμένων δεδομένων και του σημασιολογικού ιστού. Είναι μία πύλη που υποστηρίζει την παραγωγή συνδεδεμένων ανοιχτών κυβερνητικών δεδομένων συνδέοντας διάφορους καταλόγους δεδομένων από όλο τον κόσμο, αλλά κυρίως από την αμερικάνικη κυβερνητική πύλη http://data.gov . (Ding, 2011; Erickson et al, 2011) Επίσης πρέπει να επισημάνουμε την πύλη του CKAN (Comprehensive Knowledge Archive Network) που λειτουργεί σαν ένα online μητρώο για την εύρεση, ανταλλαγή και επαναχρησιμοποίηση συνόλων δεδομένων. Την στιγμή που γράφεται το κείμενο στο portal του CKAN, http://thedatahub.org, μπορεί κάποιος να βρει 3044 σύνολα δεδομένων.

Region	National / subnational regio	Name	OGD / PSI Catalogues	Coverage	Statu	Launched	Are the data "open"?
European Union	Estonia		http://pub.stat.ee/px-web.2001/Dialog/statfile1.asp	Central Government	Online		
European Union	Finland		http://data.suomi.fi/	Central Government	Online	2009	Yes
European Union	Denmark		http://digitaliser.dk/ressourcer	Central Government	Online	2009	Yes
European Union	Italy	Emilia-Romagna Open Data	http://dati.emilia-romagna.it	Region	Online	2011	Yes
European Union	Italy	Regione Piemonte	http://dati.piemonte.it	Region	Online	2009	Yes
European Union	Italy	Provincia Autonoma di Trento	http://www.territorio.provincia.tn.it/portal/server.pt?open=512&objID=862&&PageID=32157&mode=2&in_hi_userid=18720&cached=true	Province	Online	2012	Yes
European Union	Italy	Dati Italia	http://www.dati.gov.it/	National	Online	2011	No - some data are no commercial
European Union	Italy	Regione Veneto	http://dati.veneto.it	Region	Online	2012	Yes
European Union	Italy	Provincia di Roma	http://www.opendata.provincia.roma.it/	Province	Online	2012	Yes
European Union	Spain	Public Information Catalogue	http://www.proyectoaporta.es/web/guest/catalogo-de-informacion-publica	National	Beta	2010	Yes
European Union	Spain	Zaragoza Public Data Catalogue	http://www.zaragoza.es/ciudad/servicios/conjuntodatos.htm	City	Online	2010	Yes
European Union	Spain	Open Data Euskadi	http://opendata.euskadi.net/w79-home/es	Region	Online	2010	Yes
European Union	Spain	Asturias Public Data	http://risp.asturias.es/catalogo/index.html	Region	Online	2010	Yes
European Union	United Kingdom	United Kingdom	http://data.gov.uk	Central Government	Beta	January 2010	Yes
European Union	United Kingdom	Kent	http://picandmix.org.uk/categories/	Country	Online	September 2009	
European Union	United Kingdom	Northern Ireland	http://www.opendatani.info/	Region	Online	August 2009	No - License not given
European Union	United Kingdom	Lichfield	http://www.lichfielddc.gov.uk/data	District	Online	October 2009	No - Datasets cannot be used for commercial purposes.
European Union	United Kingdom	London	http://data.london.gov.uk/	Region	Online	January 2010	No - no permission is granted to reuse data.
European Union	United Kingdom	Birmingham		City	Planed		
European Union	United Kingdom	Warwickshire Open Data	http://opendata.warwickshire.gov.uk/	District	Online	2010	Yes
North America	Canada	Edmonton	http://data.edmonton.ca/	City	Online	January 2010	
North America	Canada	Nanaimo	http://www.nanaimo.ca/datafeeds/	City	Beta	July 2009	No explicit statement that data is open. Legal disclaimer states that permission must be sought.
North America	Canada	Toronto	http://www.toronto.ca/open/	City	Beta	November 2009	
North America	Canada	Vancouver	http://data.vancouver.ca/datacatalogue/	City	Beta	September 2009	Yes: Data is explicitly open
North America	Canada	Calgary		City	Planed		License not given
North America	United States of America	United States Federal Government	http://data.gov	Federal Government	Online	May 2009	Yes

North America	United States of America	California	http://www.ca.gov/data/	State	Online		
North America	United States of America	City of Ann Arbor	http://www.a2gov.org/data/Pages/default.aspx	City	Online		
North America	United States of America	District of Columbia (Washington DC)	http://data.octo.dc.gov/	City	Online	October 2008	Users must notify the OCTO and redistribute a disclaimer.
North America	United States of America	Maine	http://www.maine.gov/cgi-bin/data/index.pl	State	Online	2009	
North America	United States of America	Michigan	http://www.michigan.gov/som/0,1607,7-192-29938_54272---,00.html	State	Online		
North America	United States of America	Massachusetts	https://wiki.state.ma.us/confluence/display/data	State	Online	2010	
North America	United States of America	New York City	http://nyc.gov/html/datamine/html/home/home.shtml	City	Online	September 2009	Though not explicit, terms imply that data can be re-used.
North America	United States of America						
North America	United States of America	Portland	http://www.civicapps.org/datasets	City and local area	Beta	March 2010	
North America	United States of America	Rhode Island	http://www.ri.gov/data/	State	Online		
North America	United States of America	San Francisco	http://datasf.org/	City and County	Beta	October 2009	No explicit license or terms of use stating material can be reused.
North America	United States of America	Seattle	http://data.seattle.gov/	City	Beta	March 2010	
North America	United States of America	Utah	http://www.utah.gov/data/	State	Online		
North America	United States of America	Boston	http://www.metrobostondatacommon.org/	City			No: License not specified
North America	United States of America	Boston	http://hubmaps1.cityofboston.gov/datahub/	City			No: No permission is granted to reuse data.
North America	United States of America	Chicago	http://www.cityofchicago.org	City			No: License not specified
North America	United States of America	New Orleans	http://www.gnocdc.org/	City			No re-distribution allowed
North America	United States of America	Portland	http://portlandonline.com/	City	Planed		Not yet published.
Oceania	Australia	Australia	http://data.australia.gov.au/	Federal and State for the Gov 2 competition. Following that, Federal Govt only until further notice	Beta	October 2009	Generally CC-BY 2.5 Australia.
Oceania	Australia	New South Wales	http://data.nsw.gov.au/catalogue	State	Online	November 2009	CC-BY Licence in some cases, but mostly links to dept crown copyright statements.. as at July 2010 - Baden Appleyard
Oceania	Australia	Victoria	http://data.vic.gov.au/	State	Online	February 2010	Generally CC-BY 2.5 Australia, otherwise all rights reserved departmental copyrights
Oceania	Australia	South Australia	http://data.sa.gov.au/	State	Coming soon		State endorsed the nGILF framework,so generally CC-BY 3.0 Australia
Oceania	New Zealand	New Zealand	http://www.data.govt.nz/	Central Government	Beta	November 2009	
Russia	Russia	OpenGovData Russia Catalog	http://opengovdata.ru/	Nationwide	Online	2010	Yes
Worldwide	Worldwide	United Nations (UNdata) Catalog	http://www.undata-api.org/wiki/datasets	Worldwide	Online	2010	Yes
Worldwide	Worldwide	The World Bank Catalog	http://data.worldbank.org/data-catalog	Worldwide	Online	2010	Yes
Worldwide	Worldwide	Guardian: World government data search	http://www.guardian.co.uk/world-government-data	Worldwide	Online	2010	Yes

Πίνακας 3.6 Πίνακας πρωτοβουλιών Δ.Κ.Δ σε παγκόσμιο επίπεδο (http://opengovernmentdata.org/data/catalogues/

3.5.5. ΕΦΑΡΜΟΓΕΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΟΥΝ ΔΗΜΟΣΙΑ ΚΥΒΕΡΝΗΤΙΚΑ ΔΕΔΟΜΕΝΑ

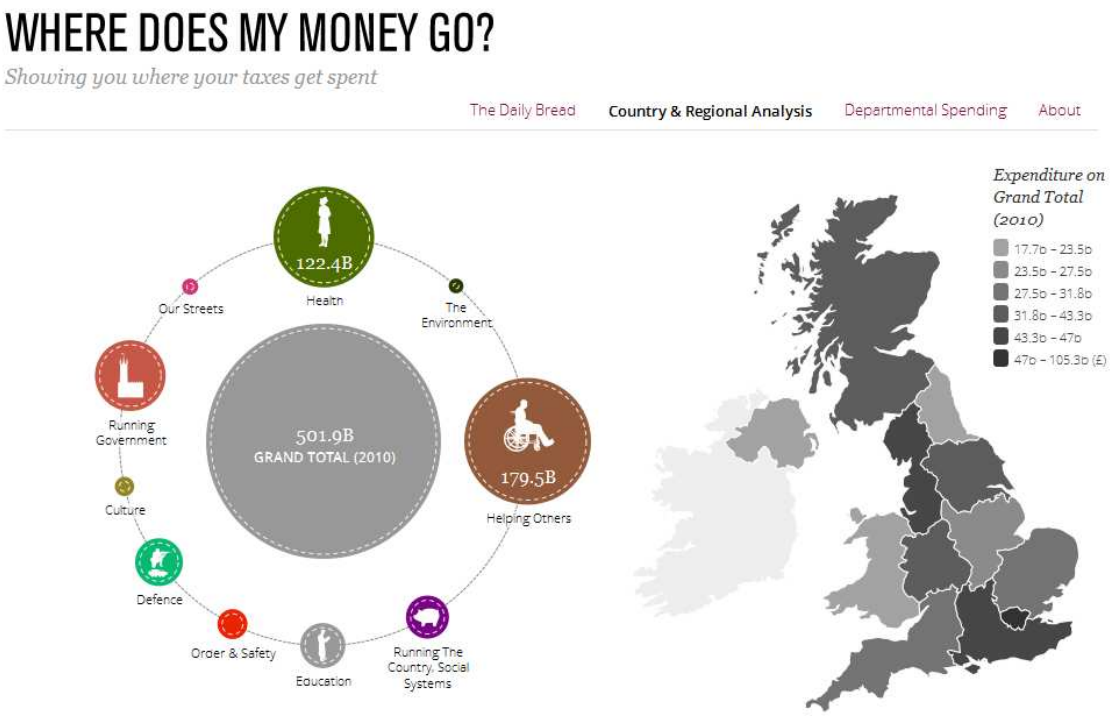
Από την στιγμή που οι κυβερνητικές πρωτοβουλίες ξεκίνησαν να «ανοίγουν» τα δεδομένα τους μέσω του διαδικτύου στο ευρύ κοινό, διάφορες εταιρείες και προγραμματιστές έσπευσαν να δημιουργήσουν εφαρμογές των οποίων το χτίσιμο βασίζεται στην επαναχρησιμοποίηση των δεδομένων αυτών. Αυτές οι εφαρμογές είναι το καλύτερο παράδειγμα για να δείξουμε την αξία της επαναχρησιμοποίησης των ανοιχτών κυβερνητικών δεδομένων.

Ένα παράδειγμα εφαρμογής είναι το Where Does My Money Go (http://wheredoesmymoneygo.org/), η οποία έχει δημιουργηθεί από το Open Knowledge Foundation και έχει ως στόχο να προωθήσει την διαφάνεια και την συμμετοχή των πολιτών μέσω της ανάλυσης και της οπτικοποίησης των πληροφοριών για τις δημόσιες δαπάνες του Ηνωμένου Βασιλείου. Οι χρήστες της υπηρεσίας είναι σε θέση να δουν που δαπανώνται ή από πού προέρχονται τα χρήματά τους, καθώς και τα χρήματα των διαφόρων υπηρεσιών και με αυτόν τον τρόπο να βοηθηθούν ώστε να ανακαλύψουν το δικό τους μέρος στην οικονομική δραστηριότητα της κυβέρνησης. (Where does my money go?, 2011)

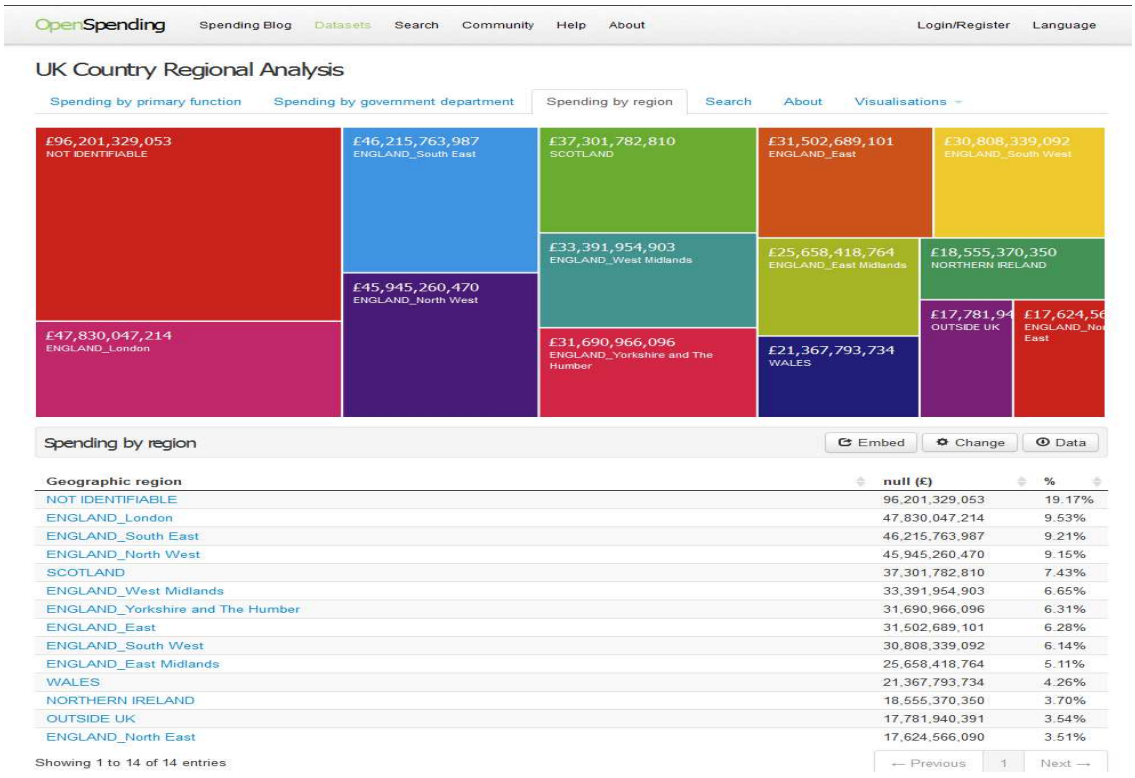
Στην εικόνα 3.7 φαίνονται οπτικοποιημένες οι δημόσιες δαπάνες του Ηνωμένου Βασιλείου ανά γενική κατηγορία π.χ. υγεία ή παιδεία, που έγιναν το 2010 καθώς και οι συνολικές δαπάνες όπως φαίνονται στην ιστοσελίδα της εφαρμογής. Τα δεδομένα που χρησιμοποιούνται στην εφαρμογή προέρχονται από την βάση δεδομένων των κυβερνητικών δαπανών του Η.Β., το COINS (Combined Online Information System). Το υπουργείο Οικονομικών, Her Majesty`s Treasury, δημοσιοποιεί τα δεδομένα αυτά μέσω του Αγγλικού κυβερνητικού portal http://data.gov.uk σε μορφή .csv αρχείων.

Ένα άλλο είδος οπτικοποίησης των δημοσίων δαπανών βλέπουμε στην εικόνα 3.8, όπου δίδεται εκτός από την εικόνα με τα ανάλογα μεγέθη ανά όγκο δαπάνης και περιοχής για το 2010 και ένας πίνακας με τα ακριβή νούμερα και την διαφορά από το προηγούμενο έτος σε ποσοστό.

Ένα άλλο παράδειγμα εφαρμογής που χρησιμοποιεί δεδομένα από την Αγγλική κυβέρνηση είναι το FixMyTransport (http://www.fixmytransport.com/). Σκοπός της εφαρμογής είναι να συγκεντρώσει αναφορές προβλημάτων για την δημόσια

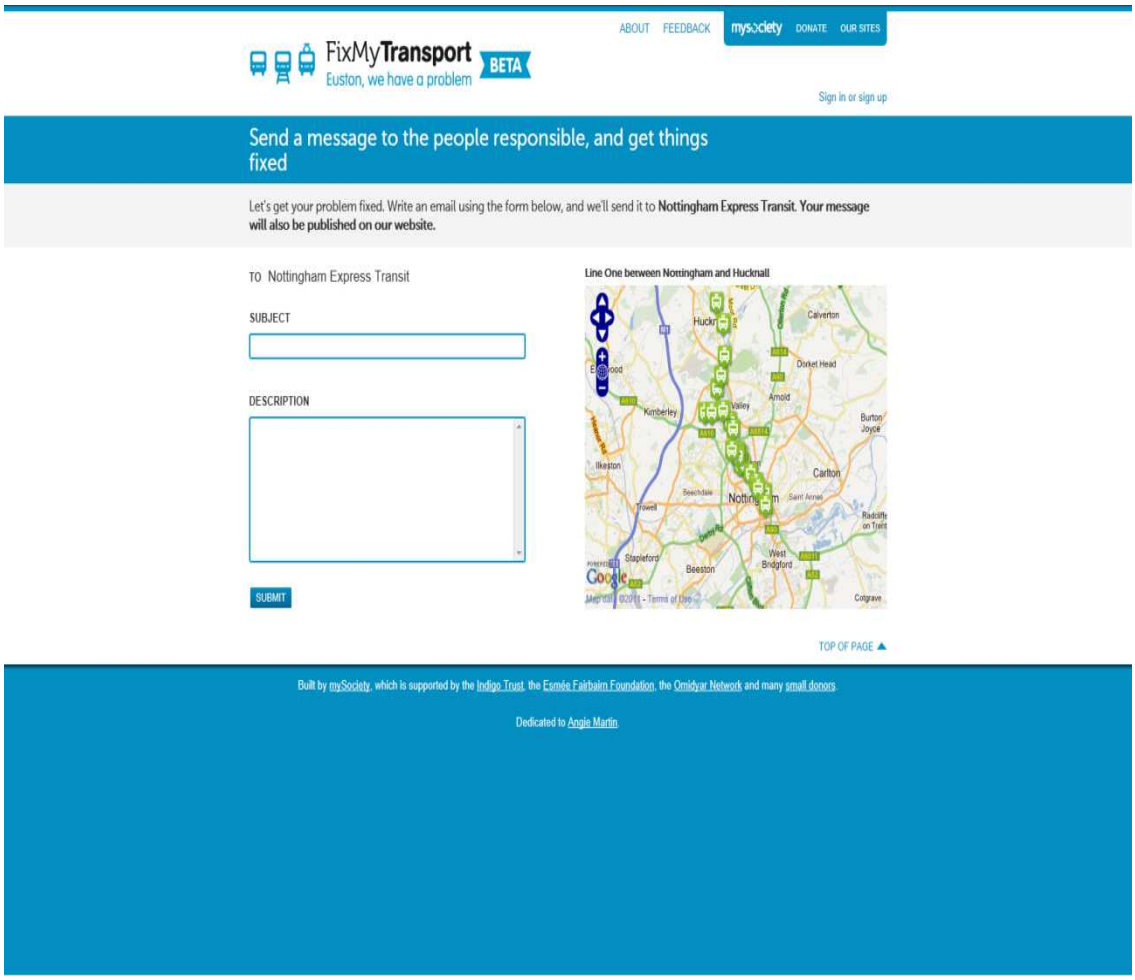


Εικόνα 3.7 Οπτικοποίηση δημοσίων δαπανών Ηνωμένου Βασιλείου ανά γενική κατηγορία (http://wheredoesmymoneygo.org/dashboard/#/~/total)



Εικόνα 3.8 Οπτικοποίηση δημοσίων δαπανών Ηνωμένου Βασιλείου ανά περιοχή (http://openspending.org/ukgov-finances-cra?_view=region)

συγκοινωνία και να τα διαβιβάσει στην κατάλληλη αρχή προς επίλυση. Η λογική είναι ότι αν ένα μικρό πρόβλημα το επισημάνουν πολλοί και διαβιβαστεί στην κατάλληλη υπηρεσία μπορεί να επιλυθεί γρηγορότερα. Προσφέρει μία απλή διεπαφή, για φυλλομετρητή από επιτραπέζιο υπολογιστή αλλά και για κινητό, για την αναφορά των προβλημάτων τα οποία μεταβιβάζονται αυτόματα στην αρμόδια υπηρεσία (Εικόνα 3.9).



Εικόνα 3.9 Διεπαφή αποστολής προβλήματος μαζί με απεικόνιση σε χάρτη (http://www.fixmytransport.com/problems/new?location_id=36348&location_type=TramMetroRoute)

Τα σύνολα δεδομένων που χρησιμοποιούνται είναι διαθέσιμα από το http://data.gov.uk και είναι τα εξής:

- National Public Transport Gazetteer (NPTG): Τοπογραφική βάση δεδομένων από το Υπουργείο Μεταφορών η οποία παρέχει ένα κοινό πλαίσιο

αναφοράς για το NaPTAN και άλλες πληροφορίες για τις δημόσιες μεταφορές

- National Public Transport Access Nodes (NaPTAN): Εθνικό σύστημα για την μοναδική αναγνώριση όλων των σημείων πρόσβασης στα δημόσια μεταφορικά μέσα
- National Public Transport Data Repository (NPTDR): Βάση δεδομένων που περιέχει ένα στιγμιότυπο από κάθε δρομολόγιο δημόσιου μέσου μεταφοράς στην Μεγάλη Βρετανία για μία επιλεγμένη εβδομάδα του Οκτωβρίου κάθε χρόνου. Τα δεδομένα είναι διαθέσιμα από το 2004 μέχρι το 2010.

Αξίζει να σημειωθεί ότι υπάρχουν πολλά παραδείγματα εφαρμογών που βασίζονται στα ανοιχτά δεδομένα των δημοσίων υπηρεσιών. Τα δύο προηγούμενα είναι αρκετά δημοφιλή και χαρακτηριστικά και αυτός είναι ο λόγος που αναφέρθηκαν αναλυτικά. Εφαρμογές που χρησιμοποιούν γεωχωρικά δεδομένα και συνδυάζουν σύνολα δεδομένων από διάφορες κατηγορίες όπως εγκληματικότητα, υγεία, δημόσιες μεταφορές και άλλες έχουν υλοποιηθεί, για φυλλομετρητές σταθερών υπολογιστών και για κινητά τηλέφωνα, και προσφέρουν πολύτιμες πληροφορίες και βοήθεια σε όσους τις χρησιμοποιούν.

3.6. ΕΜΠΟΔΙΑ ΚΑΙ ΠΡΟΒΛΗΜΑΤΑ ΣΤΗΝ ΥΛΟΠΟΙΗΣΗ ΤΩΝ ΔΗΜΟΣΙΩΝ ΚΥΒΕΡΝΗΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

3.6.1. ΕΙΣΑΓΩΓΗ

Η προσπάθεια αλλαγής νοοτροπίας κατεστημένων καταστάσεων μπορεί να φέρει αντίσταση από την πλευρά του υποβαλλόμενου στην αλλαγή. Έτσι και στις δημόσιες υπηρεσίες υπάρχουν αντιστάσεις που πρέπει να καμφθούν και απαντήσεις που πρέπει να δοθούν στις διάφορες προκλήσεις που έρχονται μαζί με την υιοθέτηση του «ανοίγματος» των κυβερνητικών δεδομένων.

3.6.2. ΠΡΟΚΛΗΣΕΙΣ

Το κίνημα για το άνοιγμα των δημοσίων κυβερνητικών δεδομένων είναι σχετικά καινούργιο και είναι στα πλαίσια της λογικής να υπάρχουν ακόμη θέματα που αφορούν την επιτυχή υλοποίηση του ανοίγματος των δεδομένων στους πολίτες και πρέπει να επιλυθούν. Θέματα όπως το νομικό πλαίσιο, η προστασία των προσωπικών δεδομένων και η σωστή προετοιμασία των δεδομένων πρέπει να επιλυθούν για να διασφαλιστεί η σωστή επαναχρησιμοποίηση τους από τους πολίτες, τις επιχειρήσεις αλλά και τις ίδιες τις δημόσιες υπηρεσίες. Οι Geiger and von Lucke (2011) επισημαίνουν τα παρακάτω έξι σημεία τα οποία χρίζουν ιδιαίτερης προσοχής:

- **Νομικό πλαίσιο:** Υπάρχουν διαφορετικοί εθνικοί και διεθνείς νόμοι για τα ανοιχτά δεδομένα, την διαφάνεια, τον έλεγχο των πληροφοριών και τις απαιτήσεις για την δημοσιοποίηση.
- **Ειδικού περιεχομένου δεδομένα:** Δεδομένα που αφορούν το δημόσιο συμφέρον, διεργασίες κυβερνητικών αποφάσεων, προσωπικά δεδομένα, πνευματική ιδιοκτησία και βιομηχανικά μυστικά πρέπει να μην δημοσιοποιούνται.
- **Η ροή και η προετοιμασία της πληροφορίας:** Υπηρεσίες και αρχές που δεν ασχολούνται ακόμη με τα ανοιχτά δεδομένα αντιμετωπίζουν μία μεγάλη πρόκληση με τα υπάρχοντα αποθηκευμένα δεδομένα τους. Πρέπει να καθοριστεί ποιά δεδομένα πρέπει να δημοσιοποιηθούν σε μορφή αναγνώσιμη από υπολογιστή και σε ποιά μορφή τελικά.
- **Φόβος παρερμηνείας και λάθος ανάγνωσης:** Αυτός που επεξεργάζεται τα δεδομένα μπορεί να τα δημοσιεύσει, αναφέροντας την πηγή τους. Υπάρχει όμως πάντα ο κίνδυνος σκόπιμης στρέβλωσης των αποτελεσμάτων της επεξεργασίας.
- **Δομή και κουλτούρα του δημόσιου τομέα:** Η αλλαγή προς τα ανοιχτά κυβερνητικά δεδομένα συνδυάζεται με αλλαγή νοοτροπίας και κουλτούρας από τον δημόσιο τομέα. Η γνώση αποτελεί δύναμη και με τα ανοιχτά δεδομένα το δημόσιο περνάει ένα μέρος της δύναμή του πίσω

στην κοινωνία. Αυτό θα μπορούσε να γίνει ζήτημα εξουσίας καθώς κάποια ξεπερασμένα μοντέλα του δημόσιου τομέα μπορεί να αντισταθούν.

- **Στρατηγικό πλαίσιο για αντιμετώπιση των στρατηγικών ελλειμμάτων:** Υπάρχει η ανάγκη ενός στρατηγικού πλαισίου για το άνοιγμα της πολιτικής και της διοίκησης, το οποίο να παρέχει κατευθυντήριες γραμμές για την προσβασιμότητα των δεδομένων, της πληροφορίας, της γνώσης, των προτύπων, των διεπαφών και του συντονισμού (Geiger and von Lucke, 2011).

Τα πλεονεκτήματα από την επαναχρησιμοποίηση των δεδομένων τα έχουμε εξετάσει στα προηγούμενα κεφάλαια, όμως υπάρχουν και μερικές αδυναμίες που πρέπει να προσεχθούν και να αντιμετωπιστούν στον δρόμο για το «άνοιγμα» των δεδομένων. Η αλλαγή νοοτροπίας των δημοσίων υπηρεσιών σε συνδυασμό με την αντίσταση των πεπαλαιωμένων μοντέλων του δημόσιου τομέα μπορεί να φέρει εμπόδια, όπως και η ανάγκη για την δημιουργία διεθνών προτύπων και η προσαρμογή των υπαρχόντων νομικών πλαισίων μπορούν να αποδειχθούν χρονοβόρες διαδικασίες.

3.6.3. Η ΠΟΙΟΤΗΤΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ Η ΧΡΗΣΗ ΤΗΣ ΩΣ ΕΜΠΟΡΕΥΜΑ

Τα δεδομένα που δημοσιεύουν οι κρατικές υπηρεσίες και γίνονται διαθέσιμα στο διαδίκτυο μπορούν να χρησιμοποιηθούν από τρίτους, αλλά και από τις ίδιες τις δημόσιες υπηρεσίες. Δεδομένα και πληροφορίες υπηρεσιών που μέχρι σήμερα χρειάζονταν χρονοβόρες διαδικασίες για περάσουν από τμήμα σε τμήμα, από την στιγμή που γίνονται «ανοιχτά» μπορούν να είναι προσβάσιμα και από άλλες υπηρεσίες.

Στις 8 Δεκεμβρίου του 2009, το Office of Management and Budget των ΗΠΑ εξέδωσε μία ντιρεκτίβα για τα Δημόσια Κυβερνητικά Δεδομένα, στην οποία έδινε εντολή στις υπηρεσίες να δημοσιοποιήσουν στο http://data.gov (την πρωτοβουλία των ΗΠΑ για τα ανοιχτά Δημόσια Δεδομένα) το λιγότερο τριών υψηλής αξίας συνόλων δεδομένων. Στην ίδια ντιρεκτίβα εξηγούσε την έννοια της υψηλής αξίας και τα δεδομένα αυτά έπρεπε να μην είχαν δημοσιοποιηθεί ξανά σε μορφή ανοιχτή και διαθέσιμη για λήψη από το διαδίκτυο. Ο Peled (2011) εξηγεί πώς η έλλειψη πλαισίου

και σαφών οδηγιών για τις δημόσιες υπηρεσίες τις οδήγησε στην τμηματοποίηση των συνόλων δεδομένων, έτσι ώστε να συμπλέουν πάντα με την εκδοθείσα οδηγία, και παράλληλα να δημοσιοποιούν τα απολύτως απαραίτητα παρά την λογική της οδηγίας, η οποία ήταν η συνεχής και αυξανόμενη δημοσιοποίηση νέων συνόλων δεδομένων από τις υπηρεσίες (McDermott, 2010).

Η έρευνα που έγινε από τον Peled (2011) οδήγησε στο συμπέρασμα ότι οι Αμερικάνικες δημόσιες υπηρεσίες δεν συνεργάζονταν, καθώς παρ` όλη την αύξηση των συνόλων δεδομένων στο data.gov από 47 σε 389.681, μόνο 5 από τις 169 υπηρεσίες που συμμετείχαν, ήταν υπεύθυνες για την δημοσιοποίηση του 99,37% του συνόλου των δεδομένων. Επίσης το 98,90% ήταν γεωγραφικές εφαρμογές οι οποίες ήταν ήδη προσβάσιμες σε ιστοσελίδα που προϋπήρχε. Το συμπέρασμα το οποίο εξάγεται από τον συγγραφέα είναι ότι οι υπηρεσίες χρησιμοποιούσαν και χρησιμοποιούν τις πληροφορίες ως εμπόρευμα για να τις ανταλλάξουν με πληροφορίες άλλων υπηρεσιών με μεγάλη αξία και το «άνοιγμα» των δεδομένων θα τις αφαιρούσε αυτό το διαπραγματευτικό πλεονέκτημα. Παράλληλα διακρίνουμε μία κατεστημένη νοοτροπία των δημοσίων υπηρεσιών η οποία αντιστέκεται στην καινούργια τάση για το άνοιγμα των δεδομένων τους.

Το πρόβλημα της ποιότητας των δεδομένων των Βρετανικών δημοσίων υπηρεσιών εξετάζει η έρευνα του EURIM (2011). Η αποφάσεις που λαμβάνονται με βάση τις πληροφορίες του δημόσιου τομέα είναι σημαντικό να στηρίζονται σε δεδομένα τα οποία είναι υψηλής ποιότητας για να αποφεύγονται ενδεχόμενα λάθη. Ένα παράδειγμα που δίνεται είναι ότι έξι εκατομμύρια Βρετανοί πολίτες φορολογήθηκαν λάθος τα τελευταία δύο χρόνια με αποτέλεσμα να χαθούν 3,8 δισεκατομμύρια λίρες, όπως και ότι 210 εκατομμύρια λίρες σε επιδόματα πληρώθηκαν τα τελευταία τρία χρόνια σε αποβιώσαντες.

Η Ελεγκτική Επιτροπή (Audit Commission, 2009), όρισε τα χαρακτηριστικά που συνθέτουν τα ποιοτικά δεδομένα:

- **Ακρίβεια:** Ακριβή για τον σκοπό που προορίζονται.
- **Εγκυρότητα:** Καταγράφονται και χρησιμοποιούνται σύμφωνα με τις σχετικές απαιτήσεις.

- **Αξιοπιστία:** Αντανακλούν σταθερές και συνεπείς διαδικασίες συλλογής δεδομένων στα σημεία συλλογής και στην πάροδο του χρόνου.
- **Επικαιρότητα:** Συλλέχθηκαν το συντομότερο δυνατόν μετά το γεγονός ή την δραστηριότητα και έγιναν διαθέσιμα μέσα σε εύλογο χρονικό διάστημα.
- **Συνάφεια:** Σχετικά με τους σκοπούς για τους οποίους πρόκειται να χρησιμοποιηθούν.
- **Πληρότητα:** Οι απαιτήσεις των δεδομένων θα πρέπει να καθορίζονται σαφώς με βάση τις πληροφορίες των αναγκών και οι διαδικασίες συλλογής των δεδομένων να ταιριάζουν με αυτές τις απαιτήσεις.

Το συμπέρασμα που μπορούμε να εξάγουμε από τα παραπάνω είναι ότι επειδή τα δεδομένα και οι πληροφορίες χρησιμοποιούνται στην λήψη σημαντικών αποφάσεων πρέπει η ποιότητά τους να είναι υψηλή. Η ποιότητα των δεδομένων παίζει σημαντικό ρόλο στο κέρδος που αυτά μπορούν να δημιουργήσουν είτε για αυτούς που τα συλλέγουν και τα δημιουργούν είτε για αυτούς που τα επαναχρησιμοποιούν. Τα παραδείγματα εσφαλμένων αποφάσεων από αλλοιωμένα δεδομένα μας εξηγούν το λόγο για την αναγκαιότητα της υψηλής ποιότητάς τους.

Συνολικά θα μπορούσαμε να συνοψίσουμε όσον άφορα τα εμπόδια και τα προβλήματα με τα λεγόμενα του Adrew Stott (μέλος του Transparency Board), ο οποίος υποστηρίζει ότι ενώ ο δημόσιος φορέας δεν βρίσκει δικαιολογίες για να μην ελευθερώσει δεδομένα, δεν το κάνει όμως γιατί τον εμποδίζουν τα παρακάτω:

1. Ανεπαρκείς πόροι.
2. Προσωπικά δεδομένα.
3. Ανεπαρκής ποιότητα δεδομένων.
4. Ανεπαρκής καθοδήγηση από την κεντρική κυβέρνηση.
5. Μικρό ενδιαφέρον για τα δεδομένα.
6. Αντίσταση του δημόσιου λόγω «ιδιοκτησίας» των δεδομένων.

(Halonen, 2012)

4. ΕΠΙΣΚΟΠΗΣΗ ΤΗΣ ΒΡΕΤΑΝΙΚΗΣ ΙΣΤΟΣΕΛΙΔΑΣ ΑΝΟΙΧΤΩΝ ΔΗΜΟΣΙΩΝ ΔΕΔΟΜΕΝΩΝ

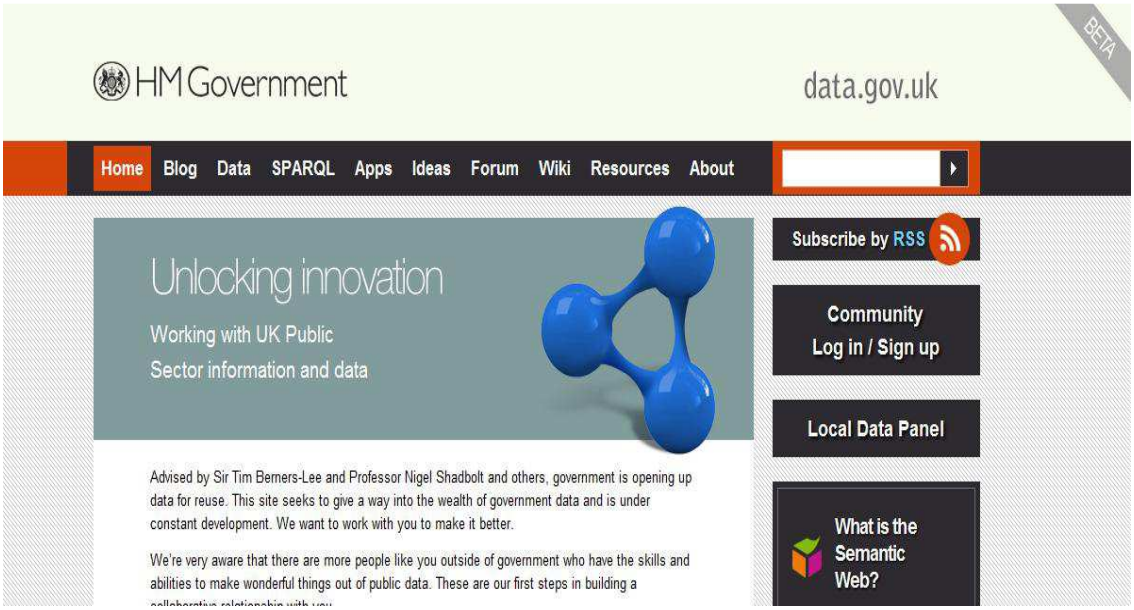
4.1. ΕΙΣΑΓΩΓΗ

Στο παρόν κεφάλαιο θα κάνουμε μία επισκόπηση της βρετανικής ιστοσελίδας των δημόσιων ανοιχτών δεδομένων προσπαθώντας να δούμε περισσότερο την δομή, το περιεχόμενο και τα σύνολα των δεδομένων που υπάρχουν ανοιχτά και διαθέσιμα για επαναχρησιμοποίηση από τους χρήστες. Ο σκοπός της εργασίας είναι η χαρτογράφηση των συνόλων δεδομένων και η εξέταση της δομής και του περιεχομένου της ιστοσελίδας γίνεται για να παρουσιάσουμε πως λειτουργεί μια κυβερνητική πρωτοβουλία για ανοιχτά σύνολα δεδομένων και σε ποια στοιχεία βασίζεται, καθώς θεωρείται η πιο ενημερωμένη και συμβατή με τις αρχές των ανοιχτών δεδομένων.

4.2. ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ

Η κλειστή beta έκδοση του http://data.gov.uk εμφανίστηκε στο διαδίκτυο στις 30 Σεπτεμβρίου 2009 με την μορφή που δείχνει η εικόνα 4.1 και μέχρι τον Ιανουάριο του 2010 πάνω από 2400 χρήστες είχαν εγγραφεί στην ιστοσελίδα για να την δοκιμάσουν, να δώσουν πληροφόρηση και να πειραματιστούν με τα δεδομένα. Κατά την διάρκεια της επίσημης πρεμιέρας της ιστοσελίδας (ανοιχτή στο κοινό beta έκδοση) τον Ιανουάριο του 2010, σ` αυτήν υπήρχαν 2500 σύνολα δεδομένων και κάποιοι προγραμματιστές είχαν φτιάξει μια ιστοσελίδα που έδειχνε τις τοποθεσίες των σχολείων ανάλογα με την αξιολόγηση του Office for Standards in Education (Data.gov.uk, 2012).

Η δημιουργία της ιστοσελίδας ήταν, την εποχή που δημιουργήθηκε, και είναι μέχρι και σήμερα κομμάτι της πολιτικής της Βρετανικής κυβέρνησης για την διαφάνεια. Ο τότε πρωθυπουργός Gordon Brown, προσέλαβε τον Sir Tim Berners-Lee και τον καθηγητή Nigel Shadbolt με σκοπό να προωθήσει την πολιτική της διαφάνειας και έτσι ξεκίνησε η ιδέα του ανοίγματος των δεδομένων που συγκεντρώνει ο δημόσιος τομέας στο ευρύ κοινό μέσω μιας ιστοσελίδας. Υπό την



Εικόνα 4.1 Beta έκδοση της ιστοσελίδας http://data.gov.uk (http://en.wikipedia.org/wiki/File:Screenshot_data.gov.uk_homepage_april_2010.jpg)

αιγίδα του Cabinet Office δημιουργήθηκε τον Ιούνιο του 2010 το Public Sector Transparency Board, το οποίο αποτελούνταν από τον πρόεδρο Francis Maude από το υπουργείο εσωτερικών, τον Sir Tim Berners-Lee διευθυντή του W3C, τον καθηγητή τεχνητής νοημοσύνης Nigel Shadbolt στο University of Southampton, τον Dr.Rufus Pollock διευθυντή του Open Knowledge Foundation και τον Andrew Stott διευθυντή του οργανισμού για την διαφάνεια και την ψηφιακή δέσμευση.

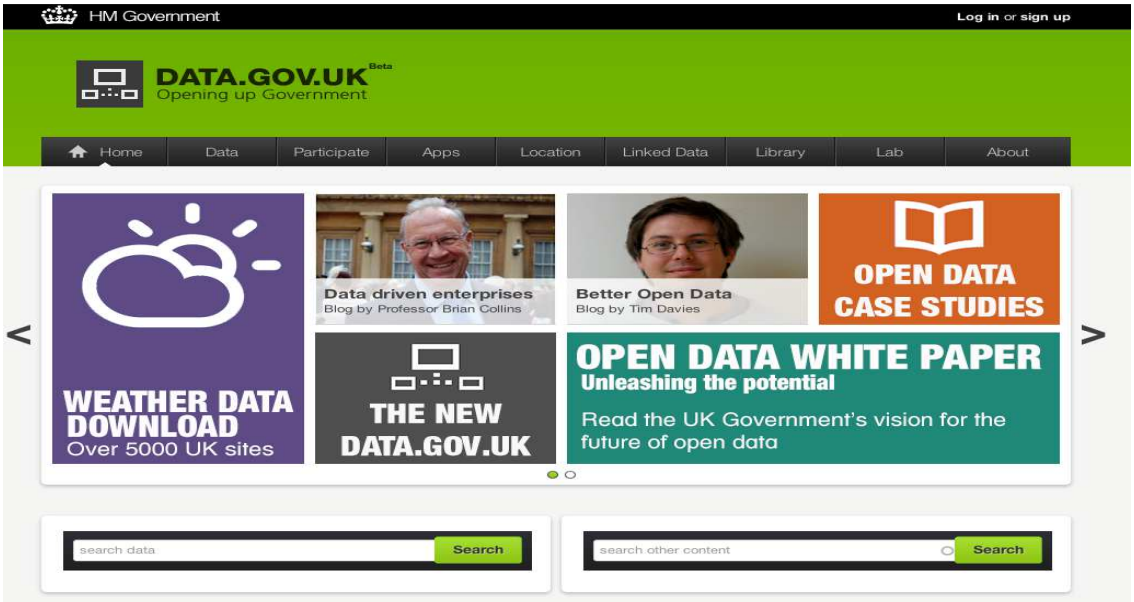
"*Government data is something we have already spent the money on... and when it is sitting there on a disk in somebody's office it is wasted.*" (Sir Tim Berners-Lee)

"Today marks an important step forward in the work the Prime Minister asked us to lead. It gives data to the public and sets the groundwork for further progress." (Professor Nigel Shadbolt)

Τα λεγόμενα των ανθρώπων που ηγήθηκαν της της προσπάθειας αυτής, όπως βλέπουμε παραπάνω, δείχνουν το νόημα του εγχειρήματος και την ουσία της κυβερνητικής προσπάθειας (Tim Berners-Lee on the UK national data website launch, 2010; Transperancy-Overview, [n.d]; The public sector transparency board-who`s who, [n.d]; Government launches one-stop shop for data, 2010; Tim Berners-Lee unveils government data project, 2010).

Η ιστοσελίδα έχει χτιστεί πάνω στην πλατφόρμα του CKAN (Comprehensive Knowledge Archive Network) που είναι ένα σύστημα διαχείρισης δεδομένων ανοιχτού κώδικα το οποίο απευθύνεται σε αυτούς που θέλουν να δημοσιεύσουν δεδομένα και υποστηρίζεται από το Open Knowledge Foundation. Από την πρώτη φορά που εμφανίστηκε η beta έκδοση στο διαδίκτυο μέχρι και σήμερα, η ιστοσελίδα έχει αλλάξει μορφή δύο φορές και η εικόνα 4.3 δείχνει την σημερινή της μορφή και η εικόνα 4.2 δείχνει την προηγούμενη της μορφή.

Εικόνα 4.2 Ιστοσελίδα http://www.data.gov.uk πριν τον Ιούνιο του 2012

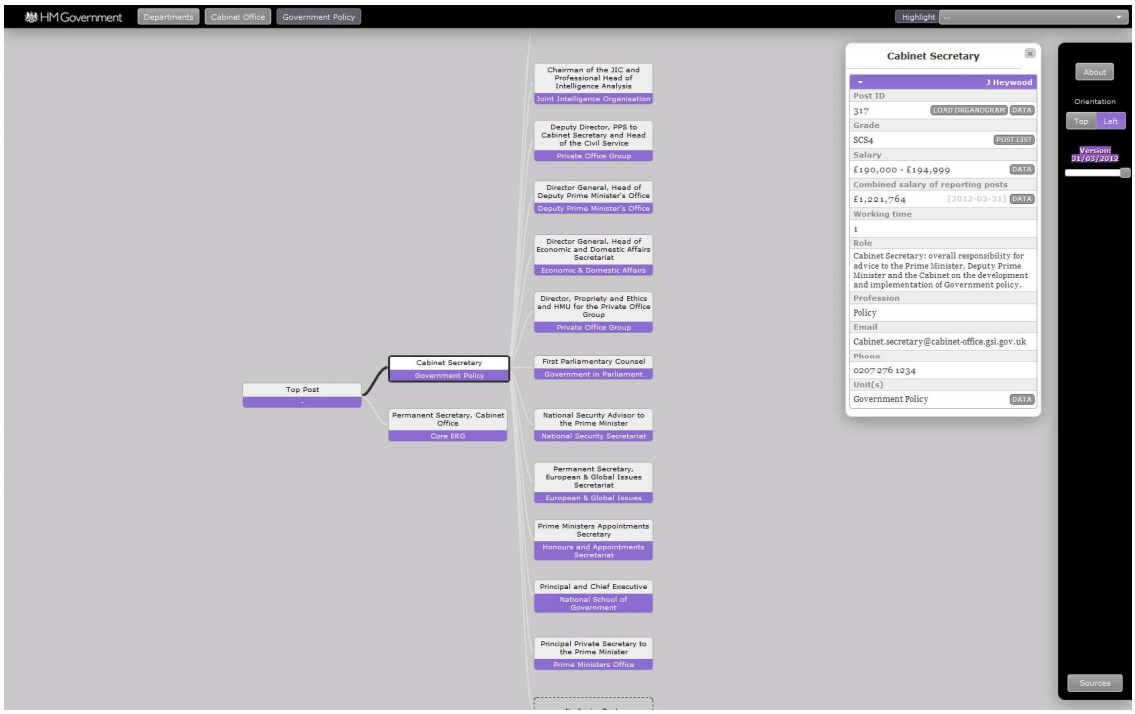


Εικόνα 4.3 Ιστοσελίδα http://www.data.gov.uk μετά τον Ιούνιο του 2012

4.3. ΔΟΜΗ ΚΑΙ ΠΕΡΙΕΧΟΜΕΝΟ

Η ιστοσελίδα λειτουργεί ως κατάλογος που συγκεντρώνει τους συνδέσμους για τα σύνολα δεδομένων τα οποία δημοσιοποιούνται από τις ιστοσελίδες των υπουργείων ή των κυβερνητικών οργανισμών και είναι διαθέσιμα προς επαναχρησιμοποίηση. Τα κυριότερα χαρακτηριστικά της δομής του http://www.data.gov.uk είναι:

1. Η σελίδα της αναζήτησης των συνόλων δεδομένων όπου υπάρχουν επιλογές αναζήτησης με κείμενο, με τον οργανισμό που είναι υπεύθυνος για την δημοσιοποίηση, με την ετικέτα και με χάρτη.
2. Η σελίδα της αναζήτησης των εφαρμογών που έχουν χτιστεί πάνω στα δεδομένα.
3. Η σελίδα που εξηγεί τα συνδεδεμένα δεδομένα.
4. Η σελίδα που έχει τον χώρο που κάποιος μπορεί να συζητήσει μέσω blog και forum με άλλους χρήστες ή με την ομάδα εργασίας της ιστοσελίδας.

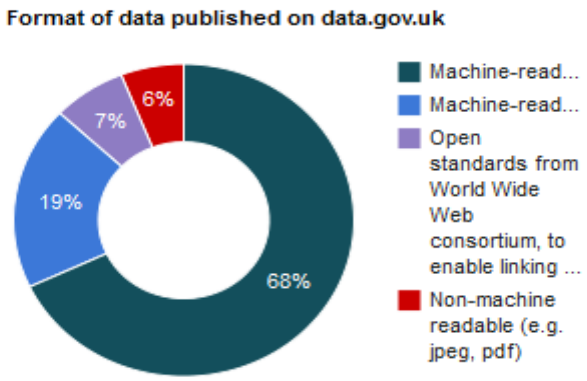


Εικόνα 4.4 Εργαλείο οπτικοποίησης οργανογραμμάτων των υπουργείων

Η λειτουργικότητα της ιστοσελίδας βελτιώθηκε με την πάροδο του χρόνου και στην τελευταία της έκδοση (εικόνα 4.3) έχουν προστεθεί κάποια στοιχεία που δεν υπήρχαν στην προηγούμενη μορφή (εικόνα 4.2). Προστέθηκε η λίστα όλων των δημόσιων οργανισμών, σε ιεραρχική μορφή, που προσφέρουν ανοιχτά τα δεδομένα τους.

Συνολικά, στην τελευταία επίσκεψη που κάναμε στην ιστοσελίδα (12/9/2012), στην ιστοσελίδα δημοσιεύουν 779 υπουργεία και οργανισμοί του δημοσίου. Επίσης βελτιώθηκε το εργαλείο οπτικοποίησης για τα οργανογράμματα και τους μισθούς των υπουργείων και των οργανισμών τους (εικόνα 4.3).

Παρόλα αυτά η ιστοσελίδα είναι λιτή γιατί η ουσία της είναι τα σύνολα δεδομένων και εκεί πέφτει όλο το βάρος. Η εικόνα 4.5 μας δείχνει τα αποτελέσματα μιας έρευνας που έκανε η ηλεκτρονική έκδοση της εφημερίδας "The Guardian", και παρουσιάστηκε στις 18 Απριλίου 2012, για το πόσο τα σύνολα δεδομένων που υπάρχουν στο http://www.data.gov.uk είναι συμβατά με τις αρχές των ανοιχτών δεδομένων όσον αφορά την μορφή των αρχείων. Τα αποτελέσματα δείχνουν ότι το 68% των συνόλων δεδομένων είναι αναγνώσιμα από ηλεκτρονικό υπολογιστή και κάτω από ανοιχτά πρότυπα (π.χ. .csv), το 19% αναγνώσιμα αλλά κάτω από κλειστά πρότυπα (π.χ. .xls), το 7% συμβατό με τα πρότυπα του W3C (π.χ. .rdf) και 6% μη αναγνώσιμα από υπολογιστή (π.χ. .pdf). Αναφερόμενοι στην έρευνα που κάναμε αξίζει να σημειώσουμε ότι τα ποσοστά των μορφών των αρχείων στις 31/10/2011 με απόκλιση μίας ποσοστιαίας μονάδας συμπίπτουν με την έρευνα του Guardian.



Εικόνα 4.5 Μορφές αρχείων (Rogers, 2012)

5. ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΡΕΥΝΑΣ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ

5.1. ΕΙΣΑΓΩΓΗ

Στο παρόν κεφάλαιο θα παρουσιάσουμε τα αποτελέσματα της έρευνας που κάναμε, με σκοπό την χαρτογράφηση των συνόλων δεδομένων του http://www.data.gov.uk , και να δείξουμε το περιεχόμενο τους και τις πιθανές σχέσεις μεταξύ τους. Θα παρουσιάσουμε επίσης τα εργαλεία που χρησιμοποιήσαμε για την επεξεργασία των δεδομένων αν και το μεγαλύτερο μέρος της έρευνας περιλάμβανε την επίσκεψη των συνδέσμων που οδηγούσαν στο κατέβασμα των αρχείων τοπικά στον υπολογιστή για την αξιολόγηση και κατηγοριοποίησή τους.

5.2. ΔΙΑΔΙΚΑΣΙΑ ΤΗΣ ΕΡΕΥΝΑΣ

5.2.1. ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΠΡΟΕΤΟΙΜΑΣΙΑ

Σύμφωνα με την μεθοδολογία που επιλέξαμε, μετά την βιβλιογραφική επισκόπηση, και σύμφωνα με τους στόχους της έρευνας τα βήματα που αρχικά σχεδιάσαμε για να ακολουθήσουμε είναι τα παρακάτω:

1. Συλλογή δεδομένων.
2. Αξιολόγηση δεδομένων.
3. Κατηγοριοποίηση δεδομένων σύμφωνα με την γενική κατηγορία.
4. Καθορισμός των σχέσεων των δεδομένων.

Οι κατηγορίες τις οποίες χρησιμοποιήσαμε για ταξινομήσουμε τα δεδομένα τις δανειστήκαμε από το Information Architecture for the UK Public Sector του UK Government Reference Architecture (Cabinet Office, 2012). Η δημιουργία του Information Architecture στοχεύει στην παροχή προτύπων και πλαισίων για να βοηθήσει στην αναμόρφωση του δημόσιου τομέα όσον αφορά την δυνατότητα των διαφόρων τμημάτων του να μοιράζονται και να επαναχρησιμοποιούν τις πληροφορίες. Έτσι η δημοσίευση και η σύνθεση των δημόσιων ανοιχτών δεδομένων και η διαχείριση της γνώσης μέσα στα πλαίσια της αρχιτεκτονικής οδηγεί σε περισσότερη επαναχρησιμοποίηση που είναι και ο στόχος της ελεύθερης

δημοσιοποίησης των ανοιχτών δημόσιων δεδομένων (Chief Technology Officer Council, 2011). Οι σχέσεις μετά την κατηγοριοποίηση αναζητήθηκαν στις ίδιες κατηγορίες και ανάλογα με το τι οι κατηγορίες περιέγραφαν με βάση τις κοινές οντότητες.

Το κομμάτι της συγκέντρωσης των απαραίτητων δεδομένων απαιτούσε την εξέταση όλων των συνόλων δεδομένων ένα προς ένα για να συλλέξουμε τα απαραίτητα στοιχεία που θα μας βοηθούσαν να ταξινομήσουμε τα σύνολα των δεδομένων κατάλληλα για να μπορέσουμε να βγάλουμε τα κατάλληλα συμπεράσματα.

5.2.2. ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

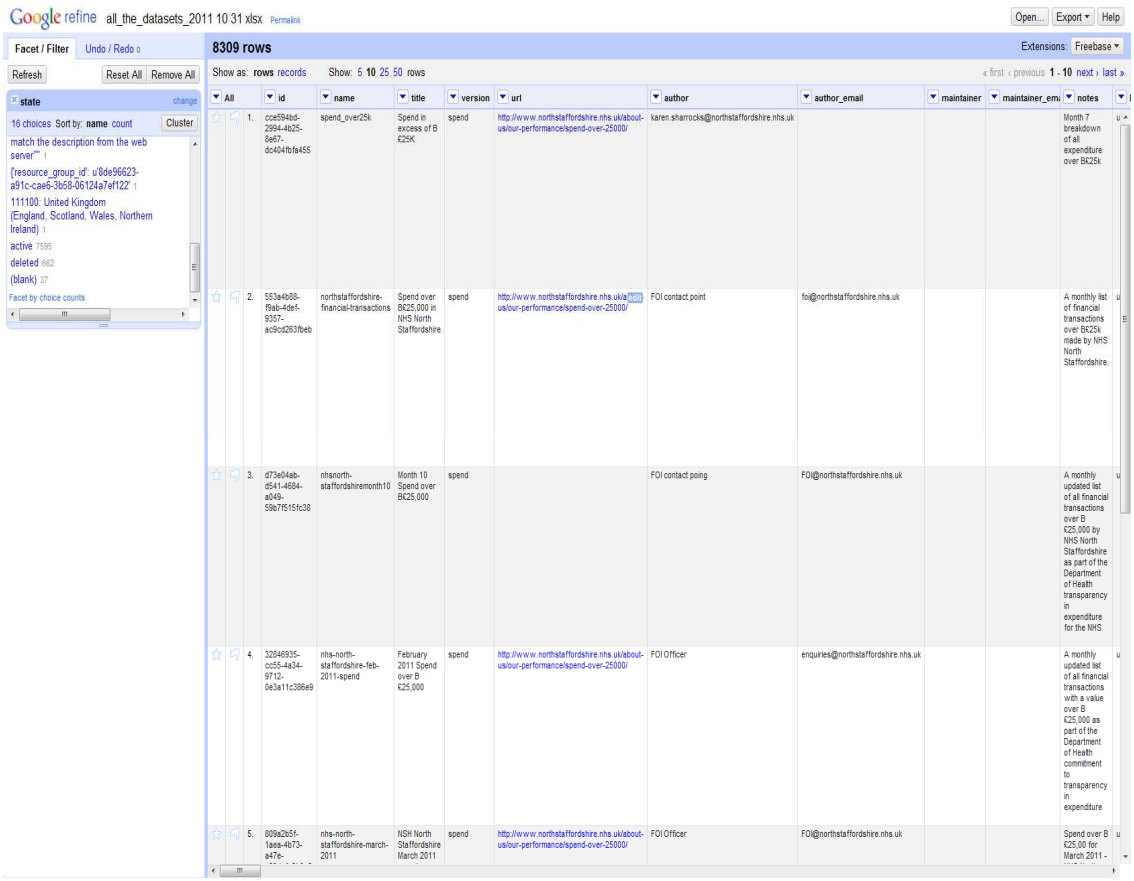
Η συλλογή των δεδομένων έγινε από την ίδια την ιστοσελίδα καθώς αυτή λειτουργεί σαν ένας κατάλογος συνδέσμων που «δείχνουν» προς τα σύνολα δεδομένων και μπορεί ο χρήστης μέσω αυτών να επαναχρησιμοποιήσει τα δεδομένα αυτά. Εκτός από τα εργαλεία που προσφέρει η ιστοσελίδα για να μπορέσει κάποιος να βρει τα δεδομένα που θέλει, εύρεση μέσω του οργανισμού που δημοσιεύει τα δεδομένα είτε μέσω κατηγοριών είτε μέσω χάρτη, υπάρχει και ένας κατάλογος αρχείων που περιέχουν μεταδεδομένα για τα σύνολα δεδομένων ο οποίος ονομάζεται dumps (http://data.gov.uk/data/dumps/archive/). Ο κατάλογος αυτός περιέχει αρχεία σε μορφή .csv και .json, τα οποία είναι στιγμιότυπα μεταδεδομένων για κάθε μέρα από την 25η Αυγούστου 2010 μέχρι και σήμερα, καθώς και για επιλεγμένες ημερομηνίες από την πρώτη ημέρα της εμφάνισης της ιστοσελίδας σε beta μορφή μέχρι και την 25η Αυγούστου.

Τα μεταδεδομένα που περιέχουν αυτά τα αρχεία δεν είναι πλήρη ώστε να μην χρειαστεί να επισκεφθεί ο χρήστης, στην περίπτωσή μας, τα ίδια τα σύνολα δεδομένων για την κατηγοριοποίηση τους. Επιλέξαμε τις παρακάτω κατηγορίες μεταδεδομένων ώστε να μπορέσουμε να συγκεντρώσουμε όλα τα σύνολα δεδομένων:

1. Το μοναδικό αναγνωριστικό του κάθε συνόλου δεδομένων (id).
2. Το URL του συνόλου δεδομένων για το http://www.data.gov.uk το οποίο μας βοήθησε να επισκεφθούμε τα σύνολα ένα προς ένα.

3. Η κατάσταση του κάθε συνόλου δεδομένων (active, deleted) για να αφαιρέσουμε τα ανενεργά σύνολα δεδομένων.
4. Οι σημειώσεις για κάθε σύνολο δεδομένων (notes) για να πάρουμε μία ιδέα του περιεχομένου των δεδομένων.
5. Οι ετικέτες για κάθε σύνολο δεδομένων (tags) οι οποίες μας βοήθησαν να κατηγοριοποιήσουμε τα δεδομένα .

Κάναμε εισαγωγή του .csv αρχείου στο Google Refine έτσι ώστε να μπορέσουμε να επεξεργαστούμε τα πεδία που μας ενδιέφεραν και να αφαιρέσουμε τα πεδία που δεν μας χρειάζονταν στην έρευνα. Το Google Refine είναι ένα εργαλείο που βοηθάει στην επεξεργασία ακατάστατων δεδομένων. Το αρχείο των μεταδεδομένων είναι ένας μεγάλος πίνακας με πολλές γραμμές και στήλες και με το εργαλείο αυτό καταφέραμε να ξεχωρίσουμε το τι έπρεπε να κρατήσουμε και τι να αφήσουμε έτσι ώστε να προχωρήσουμε παρακάτω στην έρευνα (Εικόνα 5.1).



Εικόνα 5.1 Εισαγωγή αρχείου μεταδεδομένων στο Google Refine

Στην συνέχεια με την εφαρμογή φίλτρων που διαθέτει το εργαλείο, ξεκαθαρίσαμε τα δεδομένα με αναφορά στην κατάσταση (state) του συνόλου δεδομένων αν δηλαδή το σύνολο δεδομένων ήταν ενεργό (active) ή ανενεργό (deleted). Με αυτόν τον τρόπο πήραμε έναν πίνακα ο οποίος περιείχε τα μεταδεδομένα για όλα τα ενεργά σύνολα δεδομένων την 31η Οκτωβρίου 2011 της ιστοσελίδας http://www.data.gov.uk . Με την βοήθεια του πεδίου name μπορέσαμε να επισκεφθούμε κάθε σύνολο δεδομένων ξεχωριστά έτσι ώστε να συγκρίνουμε τα μεταδεδομένα του πίνακα που είχαμε με τα στοιχεία του συνόλου δεδομένων από την ιστοσελίδα και να συμπληρώσουμε τις ελλείψεις εκεί όπου υπήρχαν. Αφαιρέσαμε τα σύνολα δεδομένων που αναφέρονταν σε γεωχωρικά δεδομένα ή περιείχαν χάρτες και φωτογραφίες καθώς δεν ήταν δυνατόν να τα κατηγοριοποιήσουμε σε γενικές κατηγορίες και δεν εξυπηρετούσαν τον σκοπό της έρευνας.

5.2.3. ΑΞΙΟΛΟΓΗΣΗ ΔΕΔΟΜΕΝΩΝ

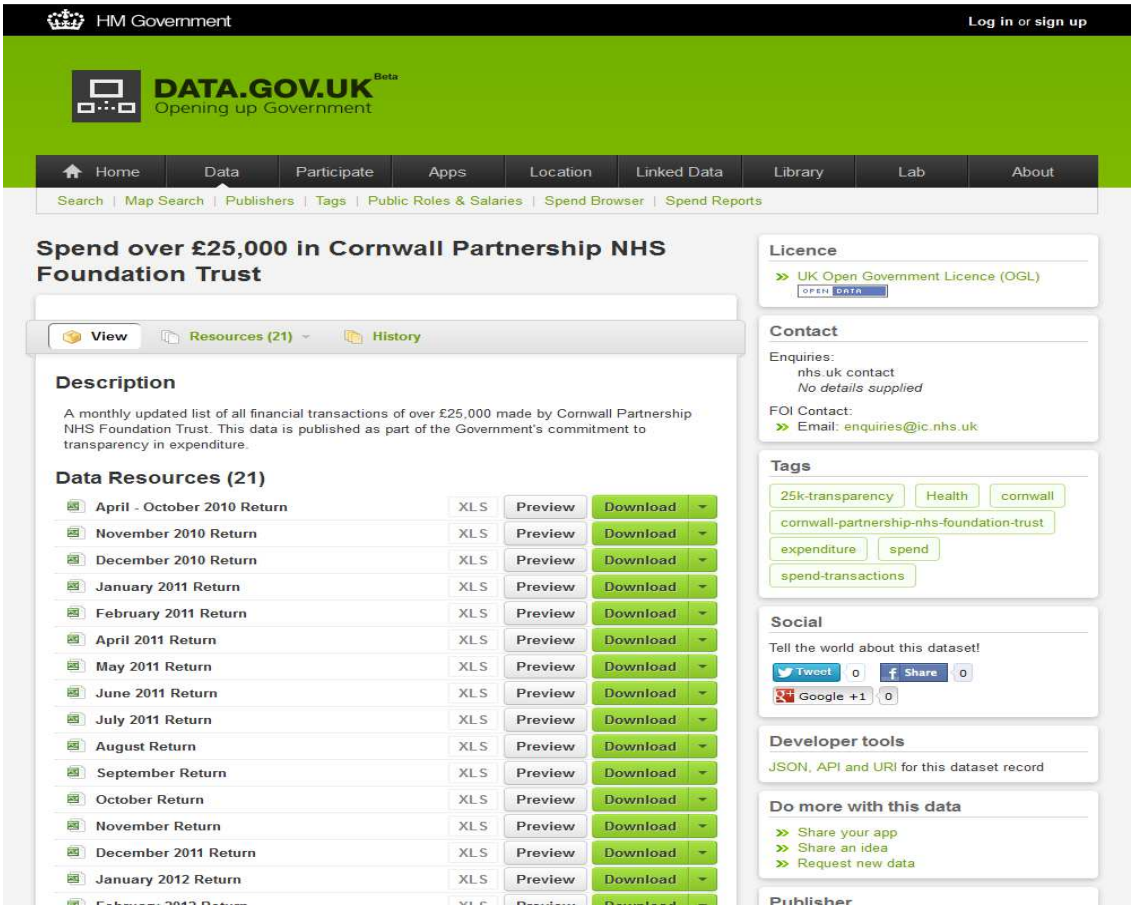
Η διαδικασία της συλλογής των δεδομένων ολοκληρώθηκε με την συγκέντρωση όλων των στοιχείων που αναζητούσαμε και τα οποία είτε λείπανε από τον πίνακα των μεταδεδομένων, είτε χρειάζονταν κάποιου είδους σύγκριση για την επιβεβαίωση τους. Η διαδικασία αυτή ήταν και η πιο χρονοβόρα γιατί χρειάστηκε να επισκεφθούμε όλους τους συνδέσμους που οδηγούσαν στα σύνολα δεδομένων περισσότερο από δύο φορές για επαλήθευση.

Στο σημείο αυτό πρέπει να κάνουμε μια διευκρίνιση όσον αφορά τον ορισμό του συνόλου δεδομένων (dataset). Σύμφωνα με τον ορισμό του διαδικτυακού λεξικού Oxford Dictionaries:

«Ένα σύνολο δεδομένων είναι μία συλλογή συνόλων πληροφοριών οι οποίες είναι σχετικές μεταξύ τους, αποτελούνται από ξεχωριστά στοιχεία και μπορεί ένας υπολογιστής να τα επεξεργαστεί ως σύνολο» (Definition of data set, 2012)

Σε συνέχεια του ορισμού και σύμφωνα με τους Renear et al. 2010, το περιεχόμενο ενός συνόλου δεδομένων μπορεί και αυτό να είναι ένα ξεχωριστό σύνολο δεδομένων δηλαδή, όπως στην περίπτωσή μας, ξεχωριστά αρχεία.

Σε πολλές περιπτώσεις ο σύνδεσμος προς ένα σύνολο δεδομένων οδηγούσε σε μία λίστα από αρχεία τα οποία αναφέρονταν στην ίδια οντότητα αλλά με διαφορετική ημερομηνία. Ένα παράδειγμα βλέπουμε στην εικόνα 5.2 όπου ο σύνδεσμος προς το σύνολο δεδομένων (http://data.gov.uk/dataset/25k_monthly_expenditure) αναφέρεται στα έξοδα άνω των 25000 λιρών του Cornwall Partnership NHS Foundation Trust, και οδηγεί σε μία λίστα από .xls αρχεία τα οποία περιέχουν τα δεδομένα για μία συγκεκριμένη χρονική περίοδο το καθένα. Σε μία άλλη περίπτωση (εικόνα 5.3) βλέπουμε ένα παράδειγμα όπου ο σύνδεσμος μας οδηγεί στο σύνολο δεδομένων (http://data.gov.uk/dataset/ni-067-percentage-of-child-protection-cases-which-were-reviewed-within-required-timescales) που αναφέρεται στο ποσοστό των υποθέσεων για την προστασία παιδιών οι οποίες εξετάστηκαν μέσα στο προβλεπόμενο χρονοδιάγραμμα και περιέχει ένα αρχείο .xls με δεδομένα από διαφορετικές χρονικές περιόδους.



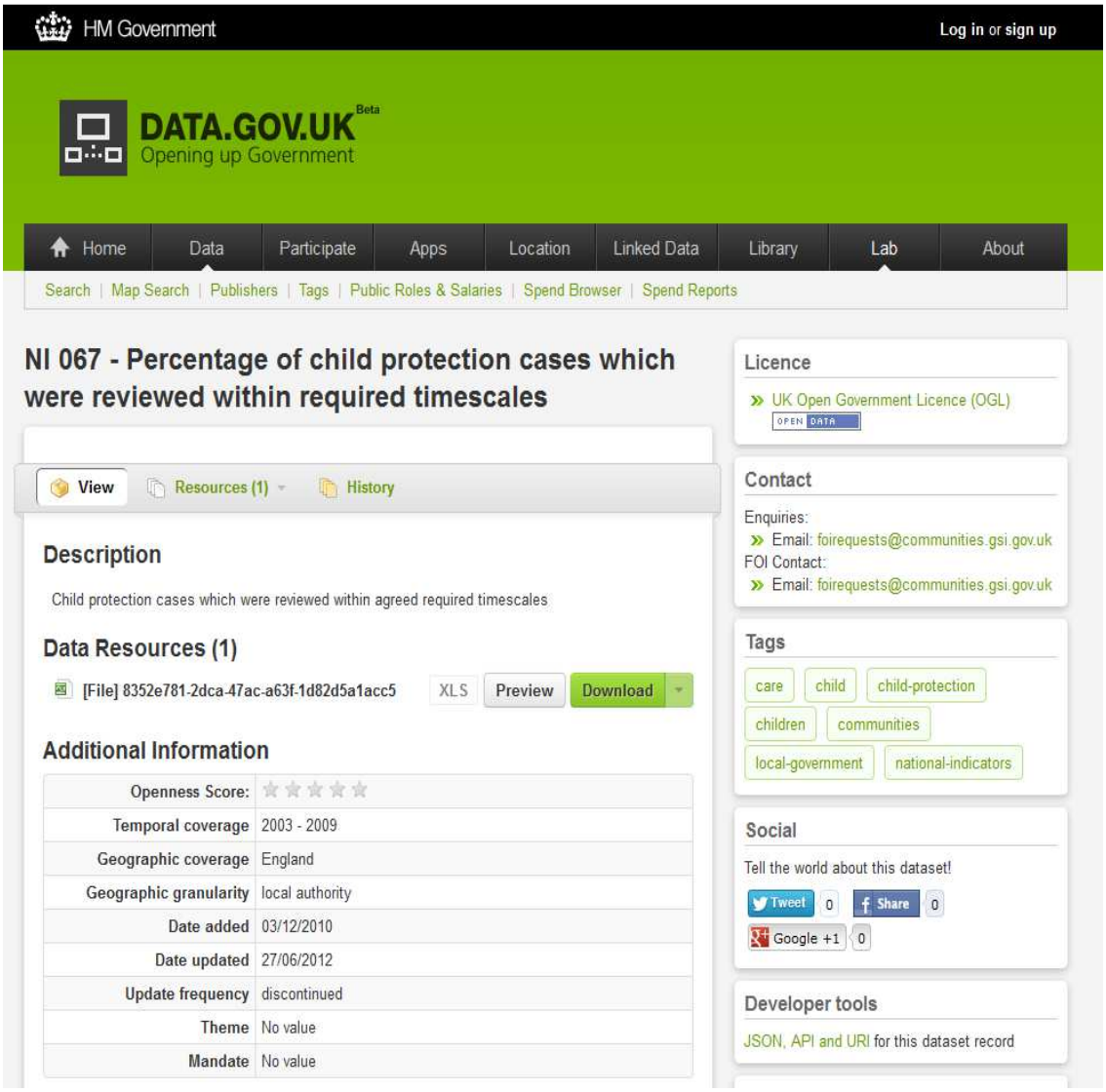
Εικόνα 5.2 Παράδειγμα συνόλου δεδομένων με πολλά αρχεία

Μετά την συλλογή των δεδομένων δημιουργήσαμε έναν πίνακα ο οποίος περιείχε τις παρακάτω στήλες για κάθε σύνολο δεδομένων:

1. Το μοναδικό αναγνωριστικό του κάθε συνόλου δεδομένων (id).
2. Το URL του συνόλου δεδομένων για το http://www.data.gov.uk .
3. Τον τίτλο (title) κάθε συνόλου δεδομένων.
4. Την γεωγραφική οντότητα την οποία περιγράφει το σύνολο δεδομένων (π.χ. England ή Scotland).
5. Την οντότητα την οποία περιγράφει το σύνολο δεδομένων (π.χ. Local Authorities ή College ή Hospital).
6. Τις σημειώσεις για την καλύτερη κατανόηση του τι περιγράφει το σύνολο δεδομένων.
7. Την γενική κατηγορία σύμφωνα με το UKRA και την ανάλογη ετικέτα (tag) των μεταδεδομένων.

Μετά από μία πρώτη αξιολόγηση των παραπάνω θεωρήσαμε ότι για να μπορέσουμε να κατηγοριοποιήσουμε καλύτερα τα σύνολα των δεδομένων έπρεπε να προσθέσουμε μία στήλη ακόμη η οποία θα περιέγραφε πιο συγκεκριμένα το αντικείμενο του συνόλου δεδομένων μετά την γενική κατηγορία (π.χ. Health/Diagnosis ή Education/Absentaism). Επίσης προσθέσαμε την χρονική περίοδο που καλύπτεται από το σύνολο δεδομένων και την εφαρμογή που πιθανόν το χρησιμοποιεί.

Μετά από την διαδικασία που περιγράψαμε προηγουμένως καταλήξαμε σε έναν πίνακα (πίνακας 5.1), που αποτελείται από 4638 γραμμές, το σύνολο των δεδομένων, και από τις στήλες που μπορούμε να διακρίνουμε στον πίνακα. Ο πίνακας 5.1 αποτελεί κομμάτι του πίνακα των συνόλων δεδομένων.



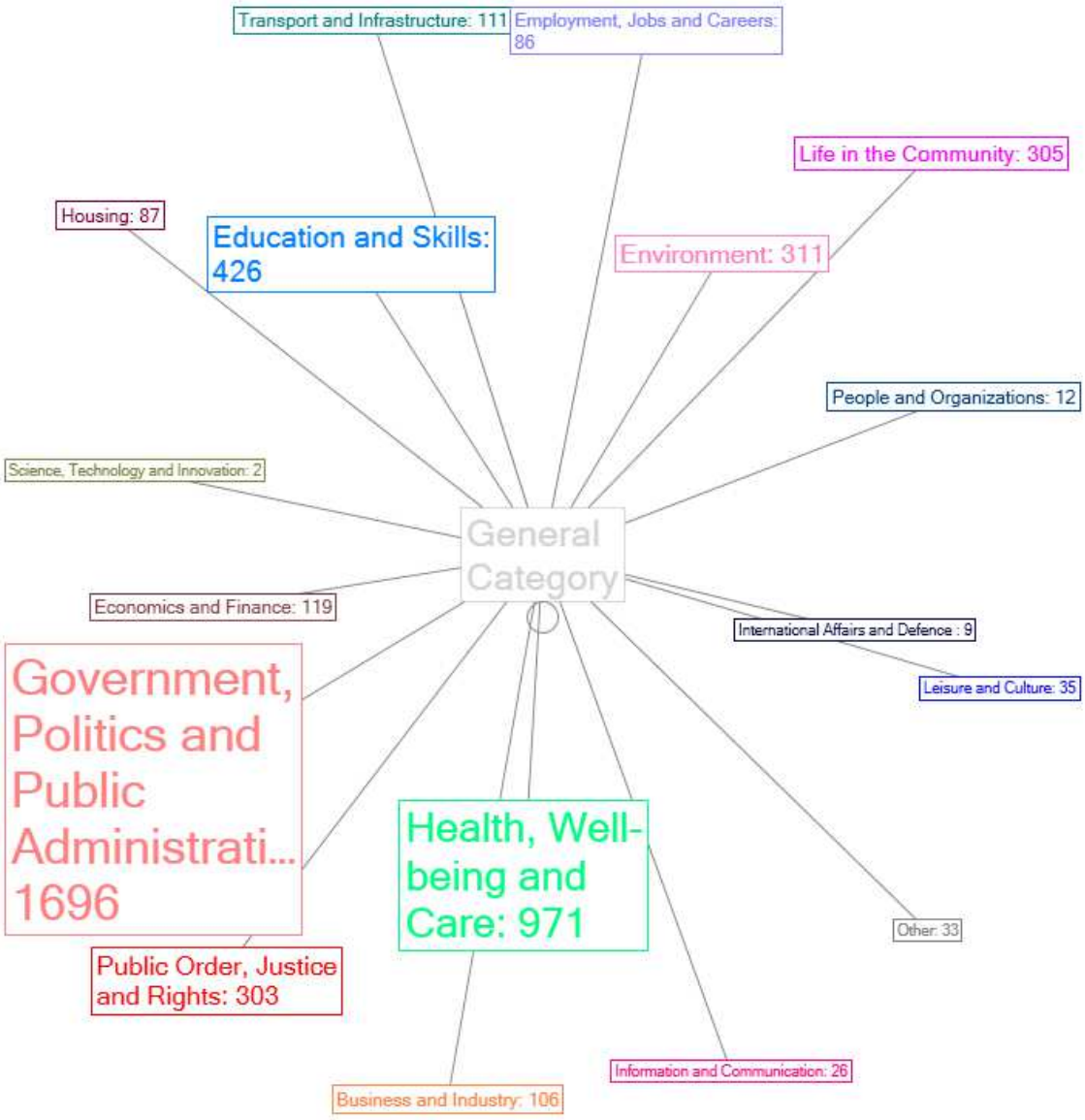
Εικόνα 5.3 Παράδειγμα συνόλου δεδομένων με ένα αρχείο

id	URL(http://data.gov.uk/dataset/...)	title	Geographical Entity	Entity	Category	Subcategory	Time coverage	related application	notes	name
47a0df62-68cd-41fe-b8ad-6c7933c25bad	accidental_dwelling_fires	Accidental dwelling fires	United Kingdom (England, Scotland, Wales, Northern Ireland)	Fire and Rescue Service (FRS) area	Health, Well-being and Care	accidental fires	2001-2009		Source: Communities and Local Government (CLG): The English	Accidental dwelling fires
9c486eaf-6d6e-475b-bb57-eb920796bc62	fire_fatalities_rate	Fire fatalities rate	England	Local Authorities	Health, Well-being and Care	accidental fires	2002-2004		deaths per 100,000 population Source: Communities and Local	Fire fatalities rate
7b280441-315a-484d-8036-fcd14f7d4c9e	green_spaces	Green spaces	England	Local Authorities	Life in the Community	Local parks	2002-2004		local parks Source: Office of the Deputy Prime	Green spaces. Percentage of residents satisfied with local parks
4d859aa6-ff14-484c-a9bc-0583d5372e8f	non-decent_local_authority_housing	Non-decent Local Authority housing	England	Local Authorities	Housing	Housing	2002-2007		decent dwellings Source: Communities and Local	Non-decent Local Authority housing
0ec9f090-18e3-4d22-b5b3-3551452d8e70	overall_employment_rate	Overall employment rate	England	Local Authorities	Employment, Jobs and Careers	Employment rate	1998-2009		active Source: Labour Force Survey (LFS)	Overall employment rate
cc1d4563-1a23-4c0e-8e51-5a95f182c356	overall_employment_rate_lfs_4_quarter_rolling_average	Overall employment rate (LFS 4 quarter rolling average)	England	Local Authorities	Employment, Jobs and Careers	Employment rate	2005-2009		Survey (ALALFS) employment rates- 4 quarter rolling average	Overall employment rate (LFS 4 quarter rolling average)
217bebdd-4719-4fd1-8922-9a68f7f37eda	vat_registrations_per_10000_adult_population	VAT registrations per 10,000 adult population	England	Local Authorities	Economics and Finance	Business support	1997-2006		population Source: Small Business Service	VAT registrations per 10,000 adult population
7556d3f1-d72a-4c37-82b5-1505e7ec53b8	employment_rate_-_ethnic_minorities	Employment rate: ethnic minorities	England	Local Authorities	Employment, Jobs and Careers	Ethnicity population	2002-2009		16+ who identify with 'non-white' ethnicity	Employment rate: ethnic minorities
0f83ca39-a3ff-4d47-be59-d27b51c89c8c	employment_rate_of_those_with_no_qualifications	Employment rate of those with no qualifications	England	Local Authorities	Employment, Jobs and Careers	No qualification population	2002-2009		qualifications	Employment rate of those with no qualifications
9228cf3b-944a-40da-af82-419c340bd7af	ni_151_-_employment_rate	NI 151 - Employment rate	England	Local Authorities	Employment, Jobs and Careers	Working age population	2004-2009		age population (16-59 for females and 16-64 for males) who are in	NI 151 - Employment rate
df868b3d-27e2-4580-8cad-83b88e1fc42d	ni_152_-_working_age_people_on_out-of-work_benefits	NI 152 - Working age people on out-of-work benefits	England	Local Authorities	Employment, Jobs and Careers	Benefits and pensions	2006-2009		percentage of the working age population who are claiming out of	NI 152 - Working age people on out-of-work benefits
7b2d2890-7b83-47e8-afb1-691fa8f97b3b	_working_age_people_claiming_out-of-work_benefits_in_the_worst_performing	claiming out-of-work benefits in the worst performing neighbourhoods.	England	Local Authorities	Employment, Jobs and Careers	Benefits and pensions	2006-2009		main out-of-work client group categories (unemployed people on	claiming out-of-work benefits in the worst performing neighbourhoods.
51578754-5608-468c-98b8-9f4cc5d4a466	s_in_an_area_showing_employment_growth	businesses in an area showing employment growth	England	Local Authorities	Employment, Jobs and Careers	Benefits and pensions	2003-2009		businesses showing year-on-year employment growth. This indicator	businesses in an area showing employment growth
81c56b3a-069b-47cb-9e13-2d6b6550663b	ni_173_flows_on_to_incapacity_benefits_from_employment	NI 173 Flows on to Incapacity Benefits from employment	England	Local Authorities	Employment, Jobs and Careers	Benefits and pensions	2005-2008		population living in a local authority who move directly from employment	NI 173 Flows on to Incapacity Benefits from employment
4dd05bd7-3b19-45cf-bfb3-90579f9d2d88	umstances_which_affect_customers_hb-ctb_benefit_entitlement_with	circumstances which affect customers HB/CTB benefit	England	Local Authorities	Health, Well-being and Care	Benefits and pensions	2009-2010		HB/CTB benefit entitlement that are processed by the local authority in a	circumstances which affect customers HB/CTB benefit
09214f0f-fd30-4089-89cc-126e9464b130	_of_people_receiving_income_based_benefits_living_in_homes_with_a_low_	people receiving income based benefits living in homes with a low	England	Local Authorities	Business and Industry	Benefits and pensions	2009		spend 10% or more of household income to maintain an adequate level	people receiving income based benefits living in homes with a low
8af06542-718a-4201-99af-8d7187a0ae84	working_age_client_group_wacg_-_claimant_rate	Working Age Client Group (WACG) - Claimant rate	England	Local Authorities	Employment, Jobs and Careers	Benefits and pensions	2005-2007		DWP claiming DWP Benefits	Working Age Client Group (WACG) - Claimant rate
0f1b1045-48ab-48db-b07f-9d73b98d5785	heart_disease_mortality	Heart disease mortality	England	Local Authorities	Health, Well-being and Care	Mortality	1996-2007		rates per 100,000 population for coronary heart disease and stroke	Heart disease mortality
de96f718-2c9f-4a94-a505-9847511da15b	ni_040_number_of_drug_users_recorded_as_being_in_effective_treatment	recorded as being in effective treatment	England	Local Authorities	Health, Well-being and Care	Drugs	2009		users (using crack and/or opiates) in treatment in a financial year, who are	recorded as being in effective treatment

Πίνακας 5.1 Κομμάτι τελικού πίνακα συνόλων δεδομένων

5.2.4. ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ

Έχοντας μία ολοκληρωμένη εικόνα από όλα τα δεδομένα που χρειαζόμασταν για να μπορέσουμε να εκπληρώσουμε τους στόχους της έρευνας, ταξινομήσαμε τα δεδομένα σύμφωνα με τις 17 κατηγορίες του Information Architecture , όπως φαίνεται στην εικόνα 5.4.



Εικόνα 5.4 Γράφημα γενικών κατηγοριών

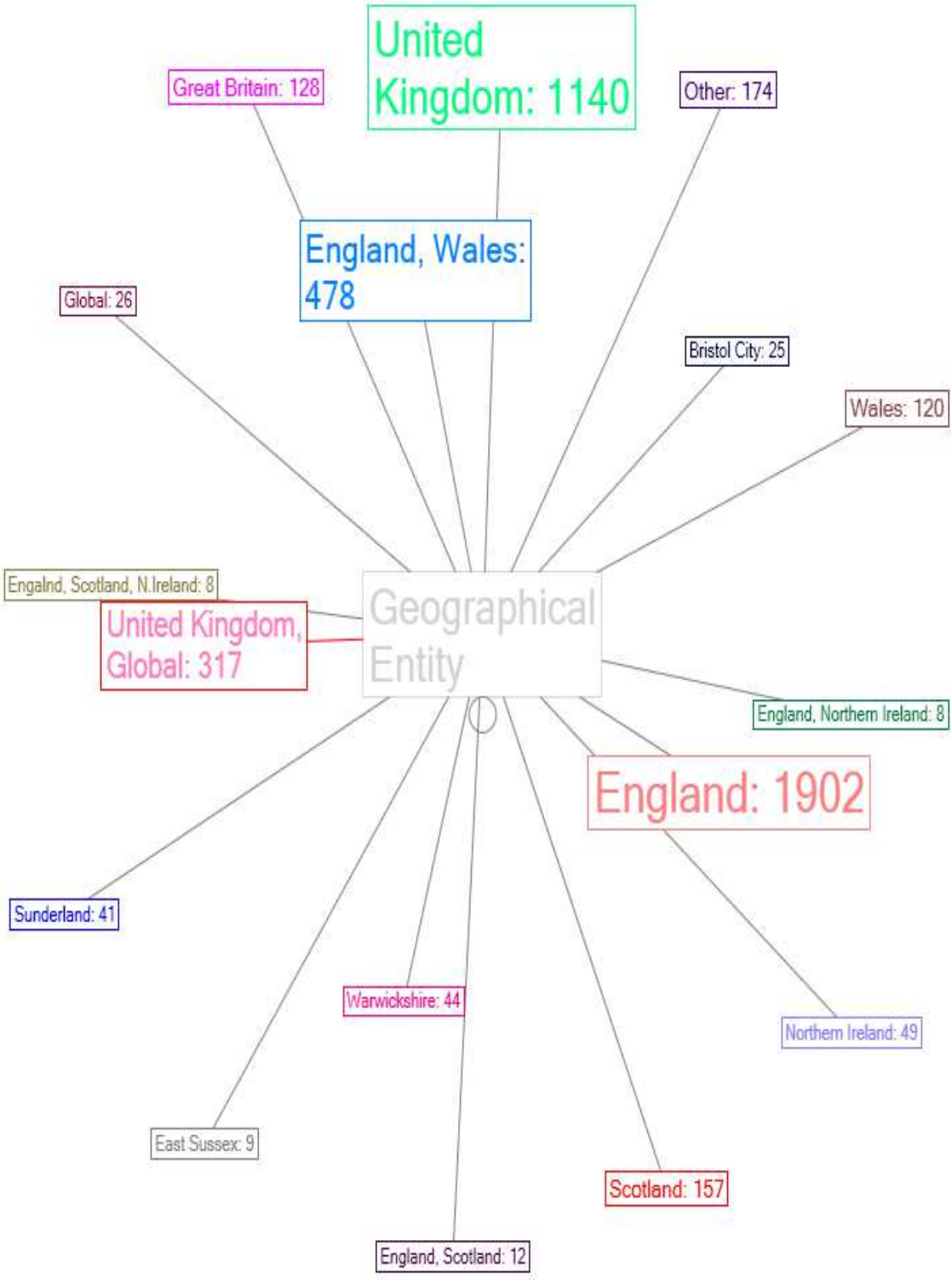
Με βάση αυτές τις κατηγορίες ταξινομήσαμε τα σύνολα δεδομένων ανάλογα με το περιεχόμενο τους και τις ετικέτες που έφεραν αυτά από τα μεταδεδομένα.

Μετά την κύρια αυτή κατηγοριοποίηση προσθέσαμε μία στήλη ακόμη στον πίνακα και την ονομάσαμε υποκατηγορία. Ο σκοπός των πεδίων της είναι να περιγράψει με περισσότερες λεπτομέρειες τι περιλαμβάνει το σύνολο δεδομένων σε σχέση πάντα με την κύρια κατηγορία.

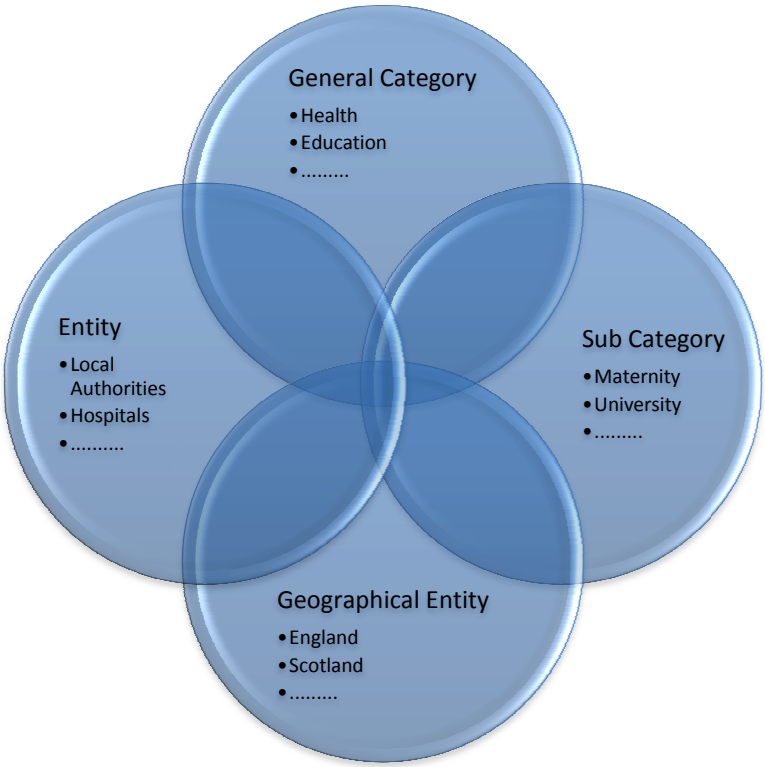
Επίσης ταξινομήσαμε τα δεδομένα σύμφωνα με την γεωγραφική οντότητα και με την οντότητα που περιγράφουν δηλαδή αν περιγράφονται νοσοκομεία ή εκπαιδευτικά ιδρύματα. Σε ιεραρχική μορφή το σύνολο των δεδομένων χωρίζεται σε γενικές κατηγορίες και γεωγραφικές οντότητες. Με την σειρά τους οι γενικές κατηγορίες έχουν τις υποκατηγορίες και οι γεωγραφικές οντότητες τις οντότητες που περιγράφουν τα δεδομένα. Το γεγονός ότι οι γεωγραφικές οντότητες και οι γενικές κατηγορίες είναι στο ίδιο επίπεδο ιεραρχίας δεν μας εμποδίζει να προχωρήσουμε σε μεγαλύτερο βάθος την ομαδοποίηση των δεδομένων. Στην εικόνα 5.4 απεικονίζεται το γράφημα των γενικών κατηγοριών των συνόλων δεδομένων και στην εικόνα 5.5 το γράφημα των γεωγραφικών οντοτήτων.

Η εικόνα 5.6 μας δίνει σχηματικά πως δημιουργείται μία ομάδα συνόλων δεδομένων με κοινά χαρακτηριστικά, καθώς κάθε πεδίο είναι και ένα χαρακτηριστικό των δεδομένων, και στο σημείο όπου τέμνονται υπάρχει απόλυτη ταύτιση των χαρακτηριστικών. Για να φτάσουμε όμως στο σημείο τομής κάνουμε ιεραρχική αναζήτηση με το εργαλείο οπτικοποίησης. Έτσι αν η αναζήτησή μας ξεκινήσει από τις γενικές κατηγορίες, το επόμενο στάδιο θα είναι οι υποκατηγορίες αυτών και στην συνέχεια μπορούμε να δούμε σε ποιές γεωγραφικές οντότητες χωρίζονται αυτές και τέλος σε ποιές οντότητες ώστε τα σύνολα δεδομένων να απόλυτα ομαδοποιημένα στο τελευταίο επίπεδο. Το ίδιο ισχύει και αν η αναζήτηση ξεκινήσει από την γεωγραφική οντότητα στην συνέχεια στην οντότητα, στην γενική κατηγορία με τελευταίο επίπεδο την υποκατηγορία.

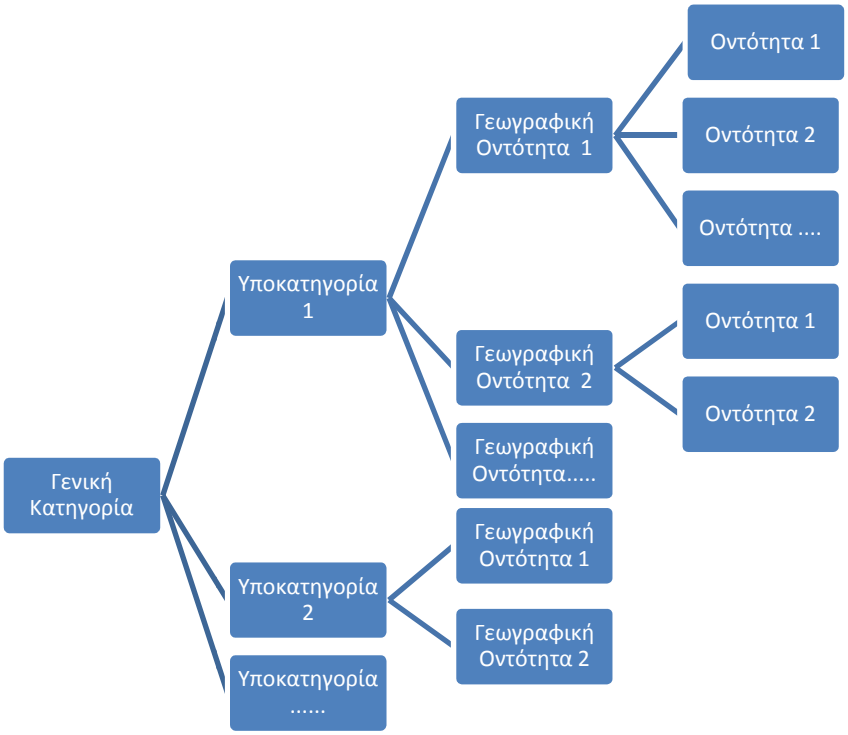
Οι εικόνες 5.7 και 5.8 δείχνουν πως ταξινομούνται ιεραρχικά οι δύο προσεγγίσεις ομαδοποίησης των δεδομένων σχηματικά, παρόλο που όπως προαναφέραμε οι γεωγραφικές οντότητες και οι γενικές κατηγορίες είναι στο ίδιο επίπεδο. Αναγνωρίζοντας τις κατηγοριοποιήσεις των δεδομένων μπορέσαμε να ξεχωρίσουμε τις σχέσεις μεταξύ των συνόλων δεδομένων, ξεκινώντας είτε από την πρώτη προσέγγιση είτε από την δεύτερη ανάλογα με το πως ο χρήστης θέλει να ομαδοποιήσει τα δεδομένα.



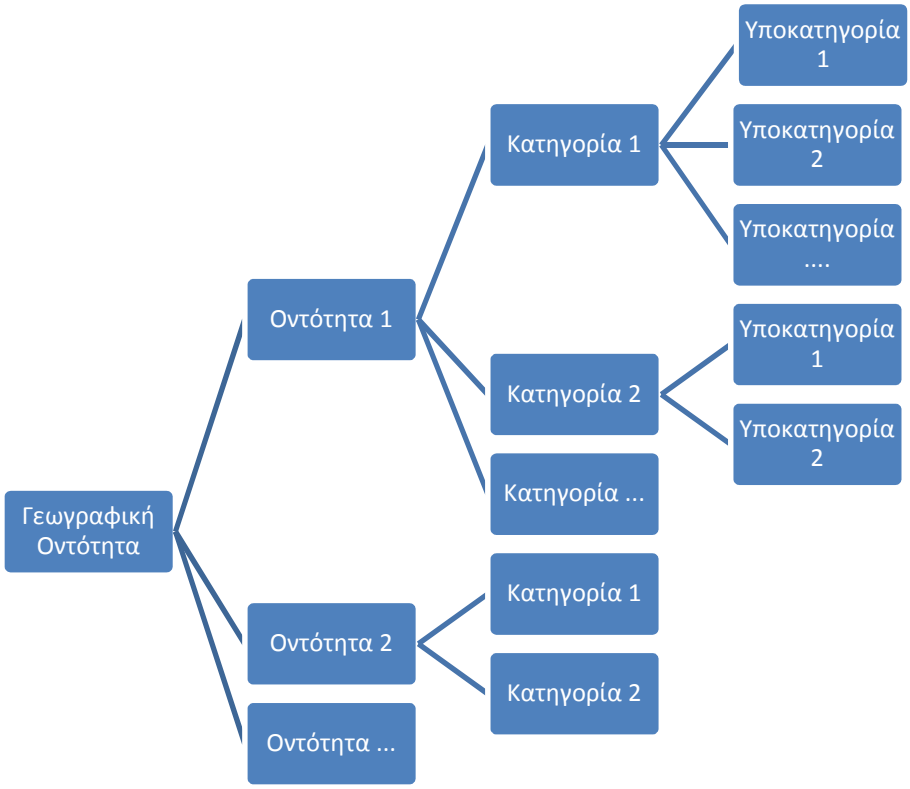
Εικόνα 5.5 Γράφημα γεωγραφικών οντοτήτων



Εικόνα 5.6 Ομαδοποίηση συνόλων δεδομένων



Εικόνα 5.7 Ιεραρχική απεικόνιση κατηγοριοποίησης των δεδομένων με εκκίνηση την γενική κατηγορία



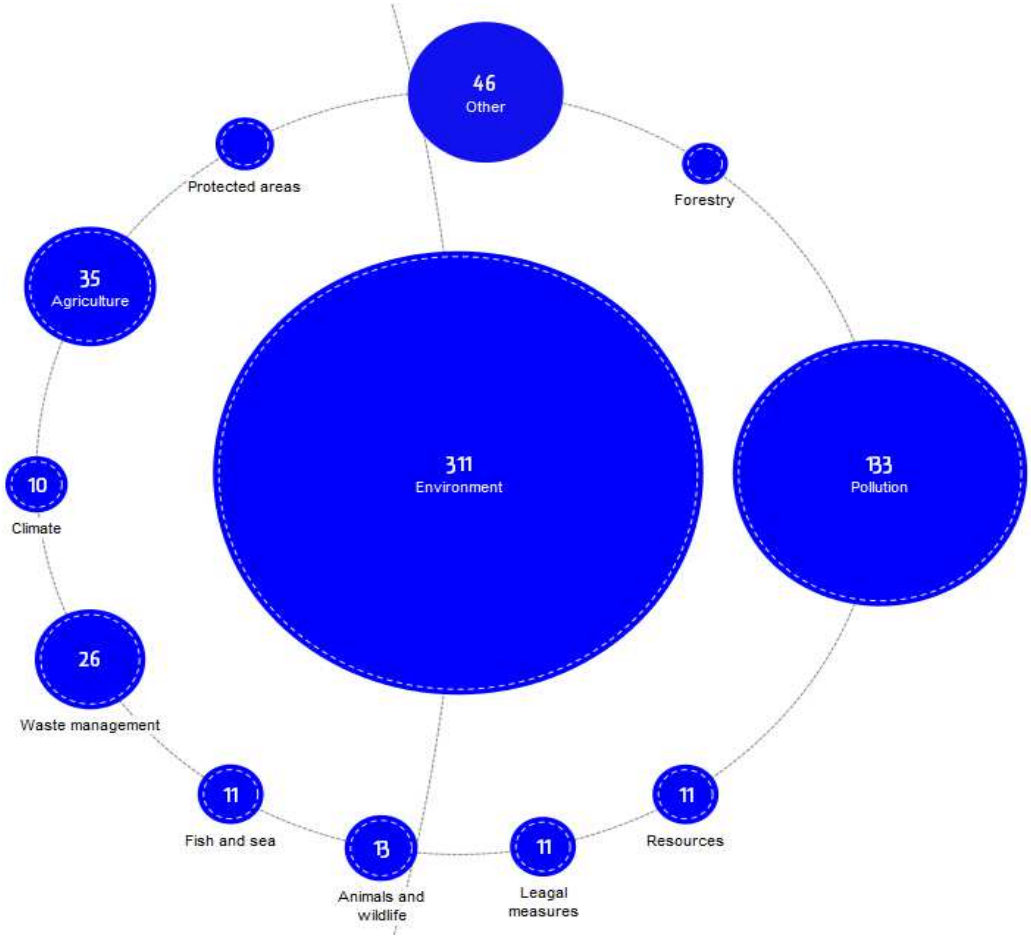
Εικόνα 5.8 Ιεραρχική απεικόνιση κατηγοριοποίησης των δεδομένων με εκκίνηση την γεωγραφική οντότητα

5.2.5. ΚΑΘΟΡΙΣΜΟΣ ΤΩΝ ΣΧΕΣΕΩΝ ΤΩΝ ΣΥΝΟΛΩΝ ΔΕΔΟΜΕΝΩΝ

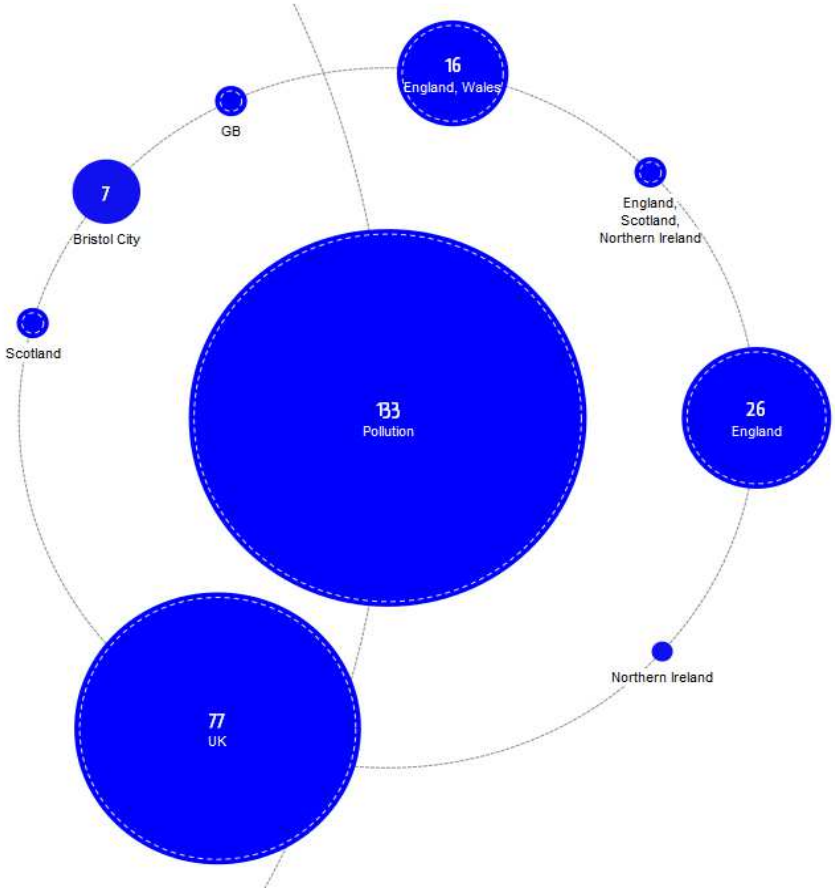
Ο καθορισμός των σχέσεων γίνεται μέσω των κατηγοριοποιήσεων που έχουμε επιλέξει και καταλήγουν σε ομάδες συνόλων δεδομένων τα οποία έχουν τα ίδια χαρακτηριστικά και μπορούν να χρησιμοποιηθούν περαιτέρω για πιθανή σύνθεση με βάση τις σχέσεις που έχουμε αναγνωρίσει.

Το αποτέλεσμα των σχέσεων αυτών είναι ομάδες συνόλων δεδομένων οι οποίες καθορίζονται από τα όμοια χαρακτηριστικά των δεδομένων και στα τέσσερα πεδία που έχουμε τιμές (εικόνα 5.6). Για να μπορέσουμε να βρούμε αυτές τις ομάδες κάνουμε ιεραρχική αναζήτηση ξεκινώντας από τα δεδομένα τα οποία είναι κάτω από μια γενική κατηγορία, π.χ. Environment, η οποία περιέχει μικρότερες ομάδες συνόλων δεδομένων με βάση την πιο ειδική κατηγοριοποίηση, π.χ. Pollution, αυτές με την σειρά τους περιέχουν δεδομένα που περιγράφουν συγκεκριμένες γεωγραφικές οντότητες, π.χ. England, και τέλος αυτές οι ομάδες δεδομένων ποιές οντότητες περιγράφουν, π.χ. Regions. Με βάση αυτή τη σειρά των σχέσεων μεταξύ

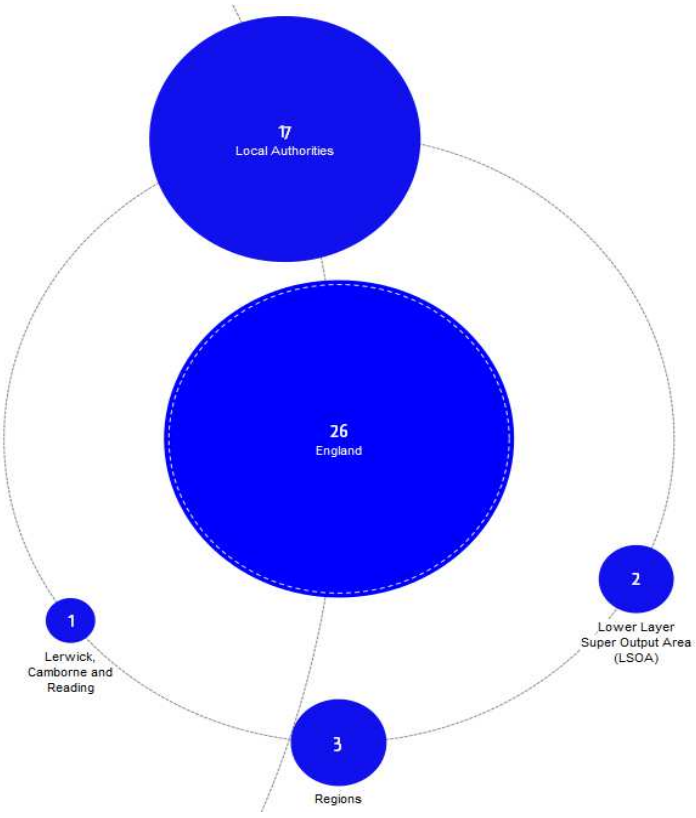
των δεδομένων φτιάξαμε την πρώτη οπτικοποίηση (εικόνες 5.9, 5.10 και 5.11) για την οποία θα μιλήσουμε σε ξεχωριστό κεφάλαιο. Μια άλλη προσέγγιση των σχέσεων των δεδομένων είναι με αφετηρία την γεωγραφική οντότητα που περιγράφει το σύνολο δεδομένων και κατάληξη την γενική κατηγορία και την υποκατηγορία την οποία αυτό περιγράφει. Ως ένα παράδειγμα αναφέρουμε μία ομάδα συνόλων δεδομένων τα οποία περιγράφουν την Αγγλία (England). Αυτή η ομάδα περιέχει υποομάδες οι οποίες κατηγοριοποιούνται με βάση την οντότητα την οποία περιγράφουν, π.χ. Local Authorities, αυτή η ομάδα δεδομένων κατηγοριοποιείται με αναφορά στις κύριες κατηγορίες, π.χ. Environment, και τέλος βρίσκουμε τις υποκατηγορίες, π.χ. Pollution. Οπτικοποίηση της συγκεκριμένης προσέγγισης βλέπουμε στις εικόνες 5.12, 5.13 και 5.14.



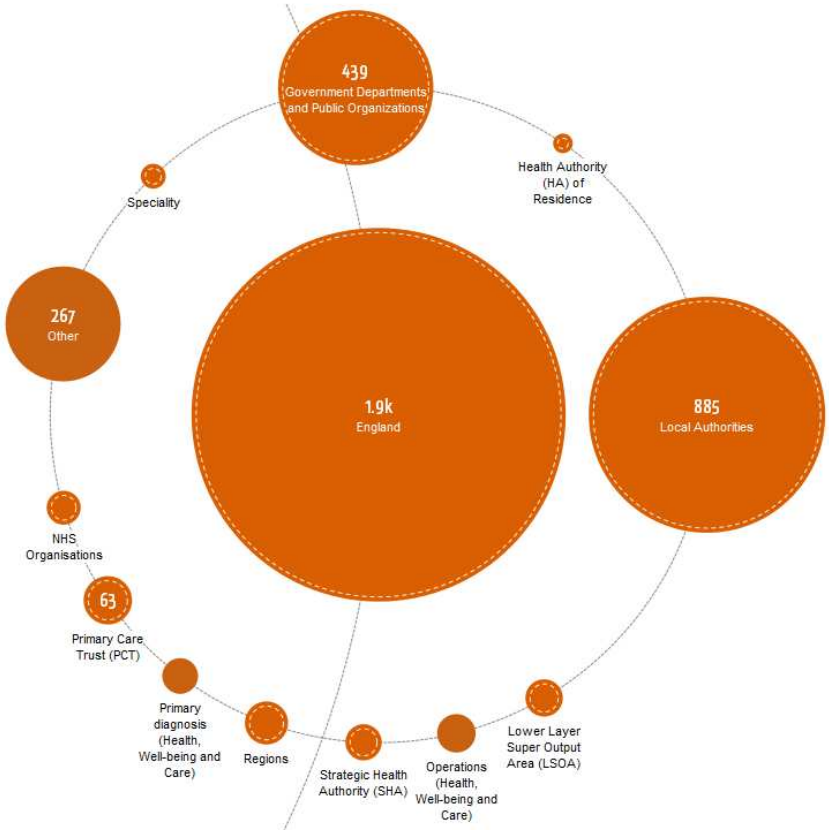
Εικόνα 5.9 Οπτικοποίηση πρώτου βήματος προσέγγισης με εκκίνηση την γενική κατηγορία



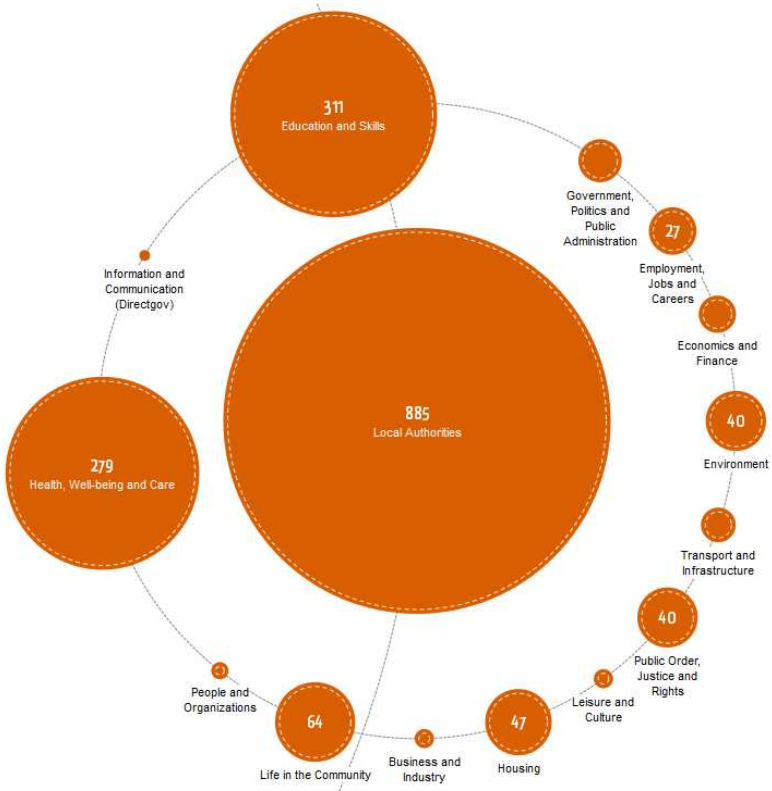
Εικόνα 5.10 Οπτικοποίηση δεύτερου βήματος προσέγγισης με εκκίνηση την γενική κατηγορία



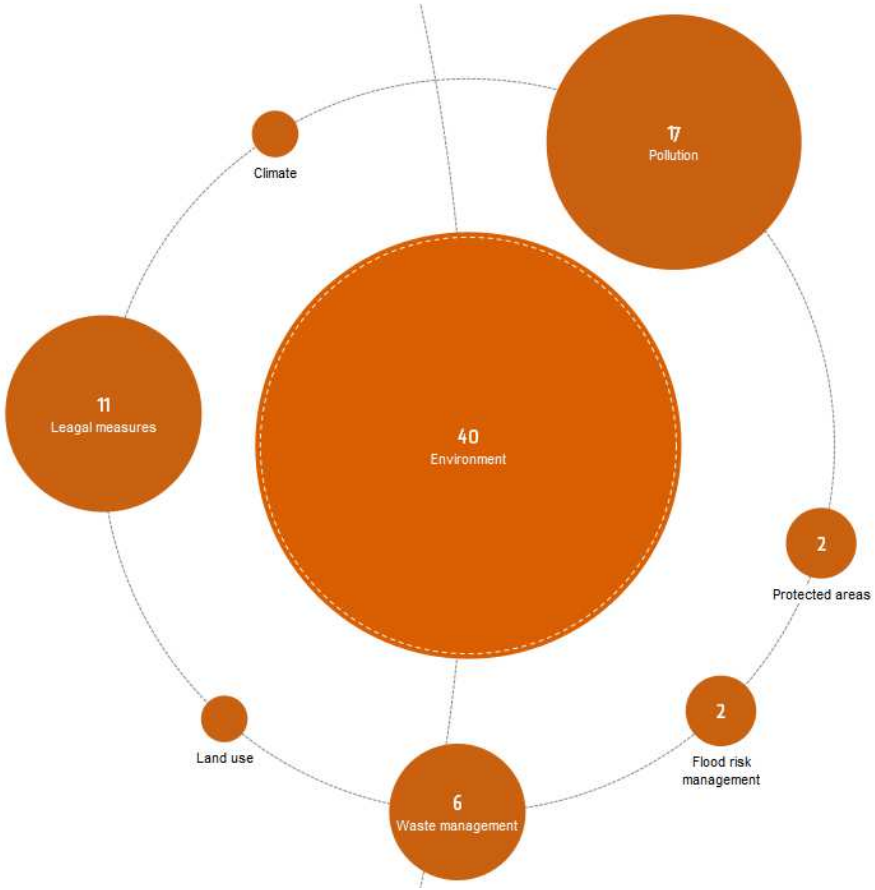
Εικόνα 5.11 Οπτικοποίηση τρίτου βήματος προσέγγισης με εκκίνηση την γενική κατηγορία



Εικόνα 5.12 Οπτικοποίηση πρώτου βήματος προσέγγισης με εκκίνηση την γεωγραφική οντότητα



Εικόνα 5.13 Οπτικοποίηση δεύτερου βήματος προσέγγισης με εκκίνηση την γεωγραφική οντότητα



Εικόνα 5.14 Οπτικοποίηση τρίτου βήματος προσέγγισης με εκκίνηση την γεωγραφική οντότητα

Με τις παραπάνω εικόνες θέλουμε να δείξουμε τον τρόπο με τον οποίο μπορούμε να ψάξουμε βαθιά στα δεδομένα και να βρούμε τα σύνολα που αναζητάμε που αναζητάμε με την ίδια βάση και έτοιμα για σύνθεση. Ένα παράδειγμα που δείχνει και τις σχέσεις των δεδομένων ακολουθεί παρακάτω. Με βάση τον τρόπο που δείξαμε προηγουμένως προχωρήσαμε στην εξής αναζήτηση:

1. Επιλέξαμε την αναζήτηση με βάση τις γεωγραφικές οντότητες.
2. Επιλέξαμε την γεωγραφική οντότητα United Kingdom.
3. Στην συνέχεια επιλέξαμε την οντότητα Local Authorities.
4. Βρήκαμε στις κατηγορίες «Employment, Jobs and Carriers» και «Economics and Finance» μία κοινή υποκατηγορία, την «Finance».

Στην συνέχεια επισκεφθήκαμε τα σύνολα δεδομένων με την χρήση του συνδέσμου που βρήκαμε στον πίνακα και κατεβάσαμε τα αρχεία. Ο πρώτος σύνδεσμος «http://data.gov.uk/dataset/annual_survey_of_hours_and_earnings» μας οδήγησε στο

σύνολο δεδομένων που έχει τον τίτλο «Annual Survey of hours and earnings» και περιγράφει το μέσο εισόδημα όλων των κατηγοριών των υπαλλήλων στο Ηνωμένο Βασίλειο από την ετήσια έρευνα του Office for National Statistics. Ο σύνδεσμος «http://data.gov.uk/dataset/personal_incomes_from_tax_records» είναι ο δεύτερος και οδηγεί στο σύνολο δεδομένων με τίτλο «Personal Incomes from tax records» και έχει τους πίνακες των μέσων εισοδημάτων και των φόρων που πληρώθηκαν για το Ηνωμένο Βασίλειο από τα επίσημα στοιχεία των φορολογικών δηλώσεων για τους υπαλλήλους και τους ελεύθερους επαγγελματίες. Ο πίνακας 5.2 είναι ένα κομμάτι του αρχείου του πρώτου συνόλου δεδομένων και ο πίνακας 5.3 είναι κομμάτι του αρχείου του δεύτερου συνόλου δεδομένων. Αυτό που μπορούμε να παρατηρήσουμε αρχικά είναι ότι στην στήλη αριστερά και στους δύο πίνακες υπάρχουν τα ονόματα των local authorities των περιοχών του Ηνωμένου Βασιλείου. Αυτό είναι το πρώτο κοινό στοιχείο πάνω στο οποίο μπορούμε να συνθέσουμε τα δύο σύνολα δεδομένων. Το δεύτερο κοινό στοιχείο είναι η στήλη του εισοδήματος των υπαλλήλων όπου περιγράφεται το μέσο ετήσιο εισόδημα ανά local authority. Το τρίτο κοινό στοιχείο είναι το έτος στο οποίο αναφέρονται τα δύο σύνολα δεδομένων, 2010, και έχοντας λοιπόν αυτά στα χέρια μας μπορούμε να παρατηρήσουμε ποιές είναι οι διαφορές ή οι ομοιότητες μεταξύ των εισοδημάτων που προέκυψαν από την έρευνα και των εισοδημάτων που μας δίνει το υπουργείο οικονομικών.

Ένα δεύτερο παράδειγμα, με εκκίνηση την γενική κατηγορία, παρουσιάζεται στα παρακάτω βήματα:

1. Επιλογή αναζήτησης με βάση την γενική κατηγορία.
2. Επιλογή κατηγορίας Environment.
3. Επιλογή υποκατηγορίας Pollution.
4. Επιλογή Γεωγραφικής οντότητας England.
5. Επιλογή οντότητας Regions.

Από τα τρία σύνολα δεδομένων με τα παραπάνω χαρακτηριστικά επιλέξαμε τον πρώτο σύνδεσμο «http://data.gov.uk/dataset/biological-quality-of-rivers-and-canals-1990-1993-to-2008» ο οποίος περιέχει δεδομένα για την βιολογική ποιότητα των ποταμών στην Αγγλία, δηλαδή για το αν υπάρχουν μικροοργανισμοί που

επηρεάζουν την ποιότητα των υδάτων. Ο δεύτερος σύνδεσμος «http://data.gov.uk/dataset/chemical-quality-of-rivers-and-canals-1990-1993-to-2008» έχει δεδομένα για την χημική ποιότητα των ποταμών της Αγγλίας, δηλαδή την περιεκτικότητα σε χημικά στοιχεία που επηρεάζουν την ποιότητα των υδάτων. Ο πίνακας 5.4 δείχνει ένα κομμάτι από τα δεδομένα του συνόλου δεδομένων για την βιολογική ποιότητα των ποταμών και ο πίνακας 5.5 τα δεδομένα για την χημική ποιότητα των ποταμών. Μπορούμε να παρατηρήσουμε τα κοινά στοιχεία των δεδομένων που είναι οι περιοχές (regions), και οι χρονολογίες αναφοράς, πάνω στα οποία μπορούμε να συνθέσουμε τα δύο σύνολα δεδομένων για να βγάλουμε χρήσιμα συμπεράσματα.

Όπως εξηγήσαμε στο πρώτο κεφάλαιο, και είναι άλλωστε και ο σκοπός της εργασίας αυτής, η ανεύρεση κοινών σημείων στα δεδομένα με σκοπό την σύνθεσή τους είναι πολύ σημαντική. Με το παρακάτω παράδειγμα προσπαθούμε να αναδείξουμε αυτήν προσπάθεια. Ξεκινώντας από τις γεωγραφικές οντότητες και συγκεκριμένα από το «UK», συνεχίζουμε επιλέγοντας την οντότητα «Local Authorities » για να μπορέσουμε να έχουμε την κοινή βάση για την σύνδεση των δεδομένων και στην συνέχεια επιλέγουμε ένα dataset «http://data.gov.uk/dataset/jobs_density» από την κατηγορία «Employment, Jobs and Careers» το οποίο μας δίνει την πυκνότητα εργασίας σε δεκαδικό αριθμό, δηλαδή τον αριθμό θέσεων εργασίας ανά κάτοικο που έχει την κατάλληλη ηλικία. Από την κατηγορία «Public Order, Justice and Rights» επιλέγουμε το dataset «http://data.gov.uk/dataset/_local-authority-recorded-crime-key-offences-2007-2009» το οποίο μας δίνει στατιστικά για τα κυριότερα αδικήματα. Η σύνθεση που θέλουμε να κάνουμε με την συγκεκριμένη επιλογή των συνόλων δεδομένων είναι να δούμε αν παίζει ρόλο η πυκνότητα εργασίας στον αριθμό των διαρρήξεων που έχουν καταγραφεί για μία συγκεκριμένη τοπική αυτοδιοίκηση, σε μία συγκεκριμένη χρονική περίοδο. Ο πίνακας 5.6 είναι η σύνθεση των δύο πινάκων των συνόλων δεδομένων με ένα επιπλέον στοιχείο που είναι ο πληθυσμός σε χιλιάδες αλλά και ο αριθμός των νοικοκυριών που υπάρχουν σε κάθε τοπική αυτοδιοίκηση ώστε να υπάρχει ακόμη ένα στοιχείο για να μπορέσουμε να εκτιμήσουμε την σχέση ανάμεσα στην εργασία και τις διαρρήξεις.

Πίνακας 5.2 Κομμάτι του πίνακα Annual Survey of hours and earnings

Table 7.7a Annual pay - Gross (£) - For all employee jobs[a]: United Kingdom, 2010

Description	Code	Number of jobs[b] (thousand)	Median	Annual percentage change	Mean	Annual percentage change	Percentiles 10	20	25	30	40	60	70	75	80	90
Cumbria	13	170	18,847	-1.5	22,427	1.4	5,722	9,701	11,514	12,854	15,830	22,999	27,752	30,964	33,688	41,921
Allerdale	289	25	17,376	0.4	20,060	4.6	x	8,668	9,552	11,571	14,538	19,488	24,529	26,910	31,426	x
Barrow-in-Furness	290	19	19,430	-4.8	21,236	-3.7	5,702	8,253	9,652	11,005	14,619	24,024	27,017	28,638	x	x
Carlisle	291	50	17,428	-0.4	20,509	3.0	5,438	9,370	10,806	12,772	15,967	20,445	25,223	26,760	29,236	x
Copeland	292	28	34,007	3.6	32,785	-0.7	9,424	14,772	18,114	20,654	29,673	36,976	40,975	42,140	43,195	x
Eden	293	17	14,997	-9.8	18,887	1.4	x	9,611	10,612	11,973	13,615	18,811	23,185	25,497	x	x
South Lakeland	294	32	17,279	-6.1	20,929	-2.9	x	10,000	11,711	12,914	15,103	21,054	25,310	27,724	x	x
Greater Manchester MC	9	878	20,277	-1.8	24,772	0.2	6,578	11,000	12,818	14,425	17,376	23,661	27,967	30,660	34,002	43,184
Bolton	207	82	17,763	-1.2	20,325	0.6	5,714	9,680	11,054	12,411	15,314	20,370	24,147	26,000	28,000	x
Bury	208	44	18,769	1.9	24,825	7.1	x	x	x	x	15,974	22,384	26,292	x	x	x
Manchester	209	231	22,416	-2.8	27,422	-3.0	7,570	12,372	14,301	16,227	19,208	26,419	31,665	34,518	38,064	49,190
Oldham	210	57	18,460	1.3	21,557	3.6	x	x	x	13,670	15,999	21,865	25,316	27,236	x	x
Rochdale	211	46	19,017	-2.0	21,433	-1.9	6,753	10,855	12,137	13,848	16,449	20,952	24,540	27,709	29,550	x
Salford	212	83	20,815	-4.6	26,019	0.5	7,361	12,418	14,008	15,096	17,341	24,179	28,004	30,270	34,598	x
Stockport	213	91	20,351	3.5	23,876	-0.5	6,517	11,024	12,977	14,725	17,583	23,501	29,349	32,416	35,685	x
Tameside	214	55	18,342	-5.7	20,433	-10.0	x	9,567	11,473	12,735	15,249	20,857	23,962	25,635	x	x
Trafford	215	105	22,596	-1.8	30,139	11.2	6,979	12,252	14,433	15,964	19,139	26,111	31,793	34,570	38,039	x
Wigan	216	86	18,477	1.5	21,734	0.3	5,944	9,416	11,007	12,662	15,817	21,725	25,155	27,611	30,022	x

Key
CV <= 5%
CV > 5% and <= 10%
CV > 10% and <= 20%
x = unreliable CV > 20% or unavailable
.. = disclosive : = not applicable - = nil or negligible

Πίνακας 5.3 Κομμάτι του πίνακα Personal Incomes from tax records

3.14 ***Income and tax by borough and district or unitary authority[(a)], 2009-10***
Taxpayers only

Government Office Region (GOR)

Numbers: thousands; Amounts: £ million; Means and Medians: £

County and district / unitary authority	*Self-employment income*			*Employment income*			*Pension income*			*Total income*			*Total tax*			
	No. of individuals	*Mean*	*Median*	*No. of individuals*	*Mean*	*Median*	*No. of individuals*	*Mean*	*Median*	*No. of individuals*	*Mean*	*Median*	*No. of individuals*	*Mean*	*Median*	*Amount*
Cumbria County	*33*	*14,700*	*9,120*	*193*	*21,700*	*17,300*	*80*	*13,100*	*11,400*	*262*	*24,000*	*18,400*	*262*	*3,490*	*2,110*	*915*
Allerdale	*6*	*12,600*	*7,860*	*35*	*22,000*	*17,100*	*14*	*13,600*	*10,600*	*47*	*23,900*	*18,300*	*47*	*3,510*	*2,020*	*164*
Barrow-in-Furness	*2*	*11,300*	*6,910*	*27*	*21,200*	*18,000*	*9*	*10,900*	*10,800*	*33*	*22,700*	*18,000*	*33*	*3,160*	*2,180*	*103*
Carlisle	*6*	*16,300*	*11,000*	*42*	*20,400*	*17,000*	*16*	*11,300*	*10,500*	*55*	*22,500*	*17,700*	*55*	*3,220*	*2,030*	*178*
Copeland	*2*	*13,100*	*8,510*	*26*	*26,900*	*22,800*	*10*	*12,400*	*12,100*	*35*	*26,300*	*21,700*	*35*	*3,870*	*2,680*	*135*
Eden	*7*	*11,700*	*6,730*	*21*	*19,900*	*16,400*	*9*	*12,900*	*10,400*	*30*	*22,700*	*17,900*	*30*	*3,150*	*2,000*	*93*
South Lakeland	*10*	*18,000*	*9,960*	*42*	*20,600*	*15,300*	*22*	*15,300*	*12,900*	*62*	*25,200*	*18,600*	*62*	*3,850*	*2,040*	*241*
Greater Manchester Metropolitan County	*110*	*18,100*	*9,450*	*973*	*21,900*	*17,700*	*271*	*12,000*	*10,700*	*1,200*	*23,900*	*18,200*	*1,200*	*3,700*	*2,180*	*4,430*
Bolton	*12*	*16,900*	*9,480*	*103*	*21,100*	*16,400*	*30*	*11,700*	*10,900*	*128*	*22,800*	*16,900*	*128*	*3,420*	*1,930*	*440*
Bury	*10*	*15,700*	*8,440*	*73*	*22,200*	*17,700*	*23*	*12,800*	*11,200*	*92*	*24,600*	*18,500*	*92*	*3,760*	*2,260*	*346*
Manchester	*15*	*16,300*	*8,020*	*160*	*20,500*	*17,000*	*24*	*10,600*	*9,890*	*182*	*22,000*	*17,100*	*182*	*3,220*	*2,010*	*588*
Oldham	*9*	*18,000*	*9,620*	*76*	*20,300*	*17,200*	*24*	*10,600*	*10,100*	*96*	*21,900*	*17,500*	*96*	*3,140*	*2,050*	*301*
Rochdale	*8*	*15,900*	*9,330*	*72*	*20,700*	*17,400*	*20*	*12,700*	*11,600*	*90*	*22,600*	*17,900*	*90*	*3,310*	*2,090*	*298*
Salford	*9*	*18,200*	*9,490*	*87*	*21,200*	*17,700*	*22*	*11,200*	*10,600*	*104*	*22,900*	*18,100*	*104*	*3,340*	*2,190*	*349*
Stockport	*17*	*21,200*	*10,800*	*110*	*25,500*	*19,900*	*37*	*14,000*	*12,300*	*142*	*28,200*	*20,800*	*142*	*4,750*	*2,670*	*677*
Tameside	*7*	*17,000*	*9,760*	*83*	*19,900*	*17,400*	*23*	*9,850*	*9,970*	*100*	*21,400*	*17,700*	*100*	*3,020*	*2,090*	*302*
Trafford	*11*	*25,800*	*9,780*	*89*	*27,300*	*19,600*	*30*	*14,100*	*12,000*	*115*	*31,000*	*20,500*	*115*	*5,870*	*2,610*	*673*
Wigan	*12*	*14,300*	*9,940*	*119*	*20,900*	*17,900*	*36*	*11,100*	*10,300*	*148*	*22,000*	*18,100*	*148*	*3,070*	*2,160*	*454*

\

Πίνακας 5.4 Κομμάτι του πίνακα Biological quality of rivers and canals

England													
			Years	Length in km							Percentage of total		
				Good		Fair		Poor	Bad		Good	Fair	Poor or
				A	B	C	D	E	F				bad
EA Region													
North West			1990	14.4	26.2	18.7	6.2	14.1	20.3		41	25	34
			1995	21.1	30.9	18.1	8.2	15.3	6.5		52	26	22
			2000	15.8	29.4	26.7	12.5	13.4	2.2		45	39	16
			2002	21.3	29.4	24	11.2	11.9	2.3		51	35	14
			2003	22.6	31.7	21.2	10.3	12.2	2		54	31	14
			2004	21.6	36.1	20.3	10.1	10.3	1.6		58	30	12
			2005	19.4	38.7	18.5	11.6	10.8	1		58	30	12
			2006	20.5	39.7	17.6	11.1	9.5	1.6		60	29	11
			2007	23.5	40.3	15.3	10.1	9.1	1.7		64	25	11
			2008	21.2	43.1	15.4	9.3	9.4	1.6		64	25	11
North East			1990	35.8	28.9	12.4	7.3	10	5.6		65	20	16
			1995	47.9	19.7	12.2	7.6	8.4	4.2		68	20	13
			2000	41.2	26.6	16.1	8.8	6.6	0.7		68	25	7
			2002	42.9	26.5	14.8	9.2	6.1	0.5		69	24	7
			2003	45.2	25.3	14.6	8.9	5.6	0.4		71	23	6
			2004	47.5	26.3	11.5	8.3	5.6	0.8		74	20	6
			2005	43.8	29.2	12.3	7.6	6.2	1		73	20	7
			2006	43.1	29.9	12.3	7.6	6.3	0.8		73	20	7
			2007	49	24.1	12.1	7.4	6.7	0.8		73	19	7
			2008	50.1	23.2	12.3	7.3	6.9	0.3		73	20	7

Πίνακας 5.5 Κομμάτι του πίνακα Chemical quality of rivers and canals

England													
			Years	Length in km							Percentage of total		
				Good		Fair		Poor	Bad		Good	Fair	Poor or
				A	B	C	D	E	F				bad
EA Region													
North West			1990	36.8	21.4	17.3	10.2	10.9	3.5		58	27	14
			1993	44.8	24.1	14.5	6.9	7.9	1.9		69	21	10
			1994	46.9	24.4	11.5	8.8	6.8	1.6		71	20	8
			1995	46.3	24.3	12.7	7.6	8	1.1		71	20	9
			1996	50.9	21.3	13.1	6.8	6.1	1.8		72	20	8
			1997	54.4	20.2	11.5	7.2	4.7	2		75	19	7
			1998	60.4	17.5	10.3	6.5	4.6	0.7		78	17	5
			1999	62.4	15.6	11.8	5.3	4.7	0.3		78	17	5
			2000	63.7	17.7	7.6	5.3	5.4	0.3		81	13	6
			2001	63.8	14.5	10.1	7.4	4.3	0		78	17	4
			2002	62.8	14.8	10.3	7.6	4.5	0		78	18	4
			2003	63.9	17.1	9.4	4.7	4.8	0.1		81	14	5
			2004	64.5	15.9	9	5.4	4.7	0.5		80	14	5
			2005	63.2	18.6	7.9	5.2	4.7	0.5		82	13	5
			2006	65.6	16.5	7.9	6.1	3.9	0.1		82	14	4
			2007	66.9	15.6	9.3	4.4	3.7	0.2		82	14	4
			2008	68.4	13.4	11.4	3	3.6	0.2		82	14	4
North East			1990	39.3	29	11.5	7.7	10.5	2		68	19	13
			1993	44.4	27.3	11.5	5	11.3	0.4		72	17	12
			1994	49.4	24.6	10.5	3.9	11.1	0.5		74	14	12
			1995	50.3	22.9	11.4	5.8	9.1	0.5		73	17	10
			1996	41.6	23.1	15.8	8.9	9.9	0.6		65	25	11
			1997	44.6	22	15.4	7.8	9.5	0.7		67	23	10
			1998	44.9	27.5	9.3	9	8.9	0.4		72	18	9
			1999	60.3	17.5	7.1	9.2	5.3	0.6		78	16	6
			2000	61.1	18.9	10.2	5.7	3.7	0.4		80	16	4
			2001	61.6	19.2	10.1	5.4	3.5	0.2		81	16	4
			2002	59.3	19.6	12.7	4.2	3.6	0.5		79	17	4
			2003	56.3	22.5	10.8	5.1	4.8	0.5		79	16	5
			2004	55.5	24.2	11	5.1	4.2	0		80	16	4
			2005	59.9	19.9	12.3	4.8	3.2	0		80	17	3
			2006	64.2	17.5	10.2	5	2.7	0.3		82	15	3
			2007	67.5	14.5	11.9	3.8	2.3	0		82	16	2
			2008	72.1	12.5	9.2	4.3	1.9	0		85	14	2

Πίνακας 5.6 Κομμάτι των πινάκων των δύο datasets, πυκνότητα εργασίας και αριθμός καταγεγραμμένων διαρρήξεων.

Local Authority	2008 Jobs density	2009 Jobs density
Allerdale	0,70	0,68
Amber Valley	0,74	0,69
Arun	0,59	0,56
Ashfield	0,68	0,68
Ashford	0,83	0,82
Aylesbury Vale	0,77	0,74
Babergh	0,72	0,68
Barking and Dagenham	0,48	0,42
Barnet	0,63	0,63
Barnsley	0,55	0,56
Barrow-in-Furness	0,70	0,76
Basildon	0,78	0,76
Basingstoke and Deane	0,85	0,82
Bassetlaw	0,71	0,70
Bath and North East Somerset	0,79	0,82
Bedford	0,77	0,76
Bexley	0,51	0,50
Birmingham	0,80	0,74
Blaby	0,90	0,89
Blackburn with Darwen	0,82	0,77
Blackpool	0,74	0,77
Blaenau Gwent	0,49	0,46
Bolsover	0,60	0,59
Bolton	0,71	0,68
Boston	0,83	0,83
Bournemouth	0,82	0,77
Bracknell Forest	0,93	0,83
Bradford	0,69	0,68
Braintree	0,64	0,65
Breckland	0,64	0,61
Brent	0,63	0,64
Brentwood	0,80	0,90
Bridgend	0,72	0,71

Local Authority	Burglary dwelling offences recorded 07/0	Burglary dwelling offences recorded 08/09	Burglary dwelling change 07/08-08/09 (%)
Allerdale	190	213	12
Amber Valley	553	381	-31
Arun	369	378	2
Ashfield	1.043	1.019	-2
Ashford	286	350	22
Aylesbury Vale	582	572	-2
Babergh	135	113	-16
Barking & Dagenham	855	1.414	65
Barnet	2.692	2.608	-3
Barnsley	952	798	-16
Barrow-in-Furness	105	132	26
Basildon	1.005	1.119	11
Basingstoke & Deane	456	664	46
Bassetlaw	609	675	11
Bath and North East Somerset	857	690	-19
Bedford	893	741	-17
Bexley	1.245	1.429	15
Birmingham	7.701	7.231	-6
Blaby	285	391	37
Blackburn with Darwen	537	575	7
Blackpool	741	758	2
Blaenau Gwent	236	234	-1
Bolsover	372	341	-8
Bolton	1.750	2.143	22
Boston	270	362	34
Bournemouth	746	801	7
Bracknell Forest	447	441	-1
Bradford	4.518	5.236	16
Braintree	268	248	-7
Breckland	121	169	40
Brent	2.556	2.535	-1
Brentwood	321	304	-5
Bridgend	402	418	4

Population figures (thousands)	Household figures (thousands)
94	42
120	53
146	67
116	51
112	47
174	71
87	38
167	68
330	135
225	98
72	32
170	74
160	68
112	49
178	76
155	66
222	93
1.010	414
93	38
141	56
142	65
69	31
74	32
262	112
58	26
163	73
114	47
497	196
141	61
130	57
270	103
72	31
134	58

5.2.6. ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΕΡΕΥΝΑΣ

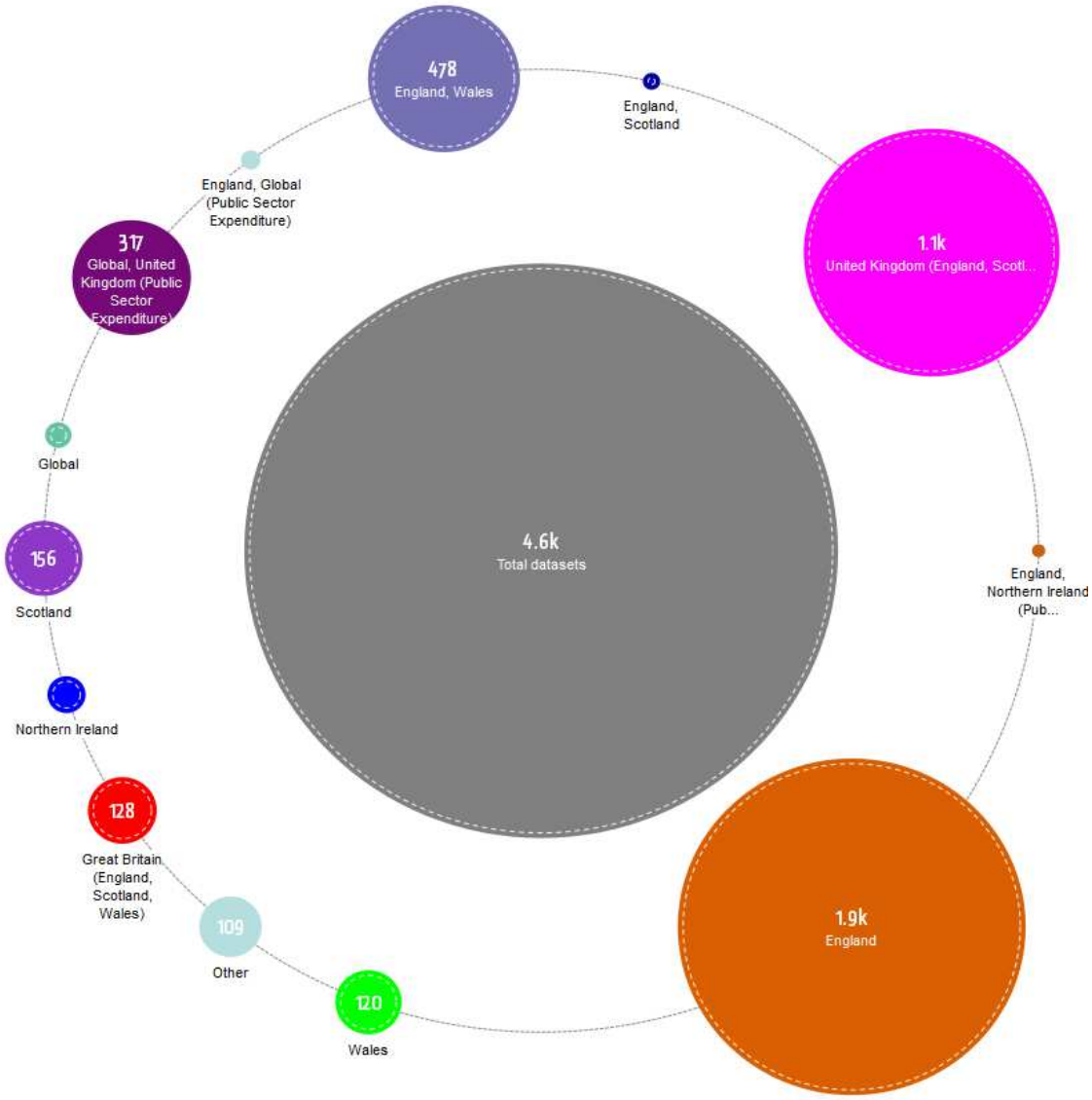
Από τις προηγούμενες ενέργειες καταλήξαμε στα αποτελέσματα τα οποία παρουσιάζουμε στους παρακάτω πίνακες. Τα αποτελέσματα παρουσιάζονται σε διαφορετικούς πίνακες, ο καθένας όμως είναι συνέχεια του άλλου. Ο πρώτος πίνακας 5.7 παρουσιάζει τον αριθμό των συνολικών δεδομένων και πως αυτά ταξινομούνται.

Σύνολα δεδομένων:		**8309**
Διαγραμμένα σύνολα δεδομένων:		714
Ενεργά σύνολα δεδομένων:		**7595**
	Γεωγραφικά σύνολα δεδομένων:	951
	Απαγορευμένης πρόσβασης:	1333
	Σύνολα που δεν βρέθηκαν:	365
	Άδεια σύνολα δεδομένων:	154
	Περιορισμένης πρόσβασης:	154
	Στατιστικά σύνολα δεδομένων:	**4638**

Πίνακας 5.7 Αριθμός συνόλων δεδομένων και γενική ταξινόμηση

Μετά την πρώτη ταξινόμηση και το ξεκαθάρισμα που κάναμε στα ενεργά σύνολα δεδομένων πήραμε τα αποτελέσματα που βλέπουμε στον πίνακα 5.7. Αφαιρέσαμε από τα ενεργά σύνολα αυτά που δεν μπορούσαμε να επεξεργαστούμε παραπάνω (γεωγραφικά, απαγορευμένης πρόσβασης, σύνολα που δεν βρέθηκαν, άδεια σύνολα και περιορισμένης πρόσβασης) και τα υπόλοιπα στατιστικά σύνολα δεδομένων τα υποβάλαμε σε περαιτέρω επεξεργασία για την κατηγοριοποίηση τους. Λαμβάνοντας υπόψη τα στοιχεία που βλέπουμε στον παραπάνω πίνακα μπορούμε να υπολογίσουμε ότι τα σύνολα δεδομένων που περιέχουν στατιστικά στοιχεία είναι το 61% του συνολικού αριθμού των δεδομένων. Όταν μιλάμε για στατιστικά σύνολα δεδομένων εννοούμε σύνολα που περιλαμβάνουν πίνακες με δεδομένα από μετρήσεις, απογραφές, συνεντεύξεις, παρατηρήσεις και άλλα παρόμοια στοιχεία.

Τα αποτελέσματα της επιπλέον επεξεργασίας μας έδωσαν τον πίνακα 5.8. Η πρώτη στήλη του πίνακα μας δείχνει τον αριθμό των συνόλων δεδομένων ανά γεωγραφική οντότητα που περιγράφει. Το Ηνωμένο Βασίλειο έχει ορισμένες διοικητικές ιδιαιτερότητες και αυτός είναι ο λόγος που βλέπουμε αυτή τη γεωγραφική διαίρεση στον πίνακα. Το επίσημο όνομα του Ηνωμένου Βασιλείου είναι «United Kingdom of Great Britain and Northern Ireland», δηλαδή έχουμε την Μεγάλη Βρετανία, η οποία αποτελείται από την Αγγλία, την Ουαλία και την Σκωτία, και την



Εικόνα 5.15 Οπτικοποίηση Γεωγραφικών οντοτήτων

Βόρεια Ιρλανδία. Παρά το γεγονός ότι οι χώρες (η Αγγλία, η Σκωτία, Η Ουαλία και η Β.Ιρλανδία) είναι διοικητικά αυτόνομες η κεντρική κυβέρνηση του ΗΒ ασκεί πολιτική για θέματα όπως άμυνα ή εξωτερική πολιτική.

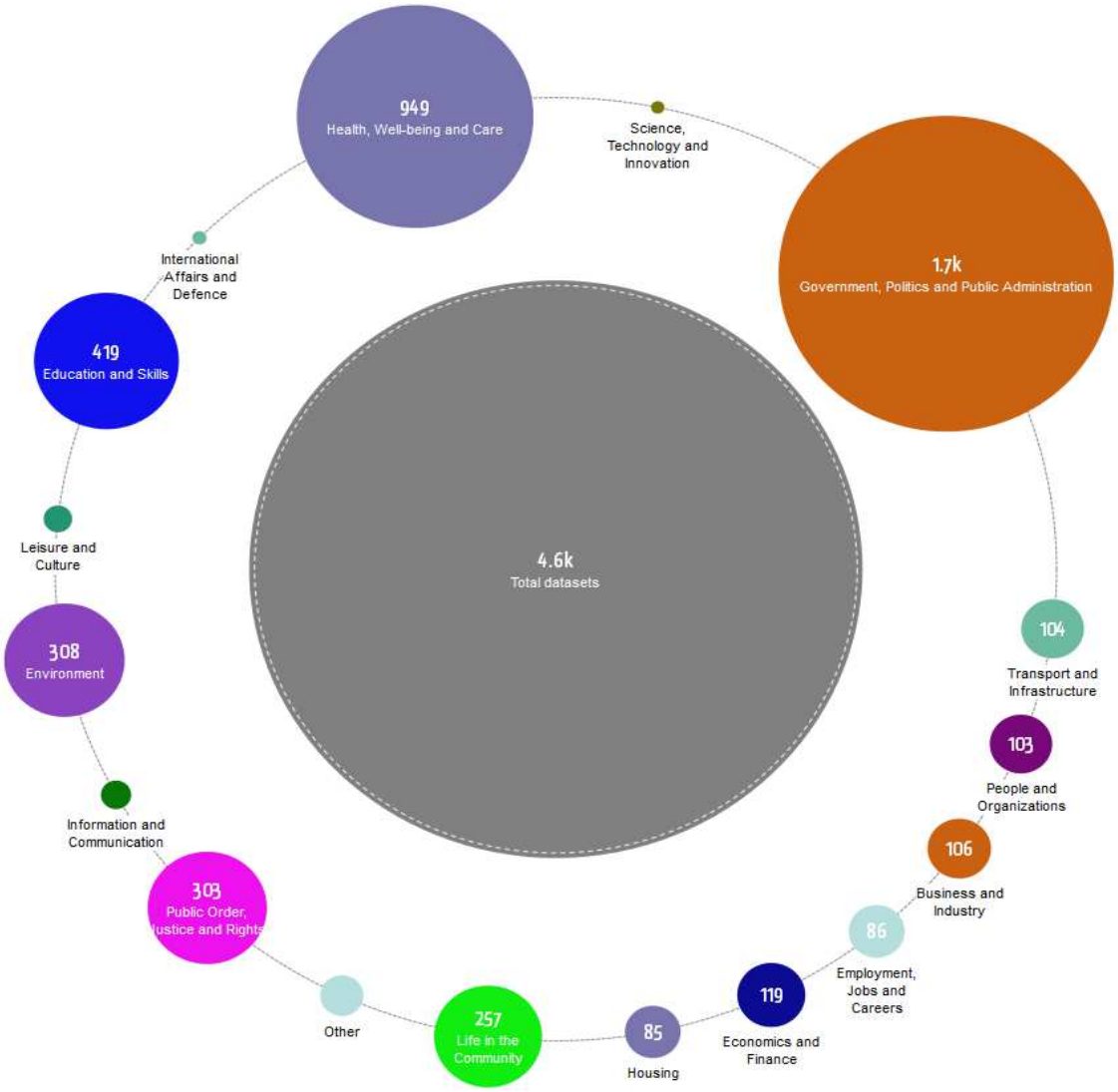
Όλες οι χώρες έχουν διαρκή παρουσία στο κοινοβούλιο (The British Government, 2011). Η δεύτερη στήλη μας δείχνει τον αριθμό των συνόλων δεδομένων ανά γενική κατηγορία όπως έχουμε εξηγήσει σε προηγούμενη αναφορά μας.

Στατιστικά σύνολα δεδομένων: 4638			
Αριθμός ανά γεωγραφική οντότητα		Αριθμός ανά γενική κατηγορία	
England:	1902	Government, Politics and Public Administration:	1696
United Kingdom:	1140	Health, Well-being and Care:	949
England, Wales:	478	Education and Skills:	419
United Kingdom, Global:	317	Environment:	308
Scotland:	157	Public Order, Justice and Rights:	303
Great Britain:	128	Life in the Community:	257
Wales:	120	Economics and Finance:	119
Northern Ireland:	49	Business and Industry:	106
Global:	26	Transport and Infrastructure:	104
Warwickshire:	44	People and Organizations:	103
Sunderland:	41	Employment, Jobs and Careers:	86
Bristol City:	25	Housing:	85
East Sussex:	9	Information and Communication:	31
England, Scotland, N.I. :	8	Leisure and Culture:	28
England, Scotland:	12	International Affairs and Defence:	9
England, Northern Ireland:	8	Science, Technology and Innovation:	2
Other:	174	Other:	33

Πίνακας 5.8 Αριθμός συνόλων δεδομένων και ταξινόμηση ανά κατηγορία

Τα περισσότερα σύνολα δεδομένων, όσον αφορά την γεωγραφική οντότητα που περιγράφουν, αφορούν την Αγγλία και είναι το 41% των στατιστικών δεδομένων. Ακολουθούν αυτά που περιγράφουν το Ηνωμένο Βασίλειο και αποτελούν περίπου το 25% του συνολικού αριθμού ενώ η Σκωτία, η Ουαλία και η Β.Ιρλανδία έχουν πολύ λιγότερα σύνολα δεδομένων που τις περιγράφουν αποκλειστικά σαν χώρες και είναι 3,5%, 2,5% και 1% αντίστοιχα. Φυσικά στον πίνακα βλέπουμε την Ουαλία να περιγράφεται στο 10% των συνόλων δεδομένων αλλά σε συνδυασμό με την Αγγλία. Ακόμη μπορούμε να παρατηρήσουμε ότι έχουμε δεδομένα που περιγράφουν αποκλειστικά κάποιες πόλεις ή περιοχές της Αγγλίας όπως το Sunderland, το Bristol City, το Warwickshire, το East Sussex και ο αριθμός τους σε σχέση με το σύνολο είναι αρκετά σεβαστός. Υπάρχουν και άλλες 87 συγκεκριμένες περιοχές που περιγράφουν τα δεδομένα οι οποίες έχουν από ένα σύνολο, οι περισσότερες, έως έξι και περιλαμβάνονται στον πίνακα στο πεδίο other. Η κάθε γεωγραφική οντότητα κατηγοριοποιείται σε οντότητες που περιγράφουν τα σύνολα δεδομένων και η κάθε κατηγορία σε πιο ειδικές υπό-κατηγορίες που αναφέρονται στην μεγαλύτερη κατηγορία.

Η δεξιά στήλη του πίνακα 5.8 έχει τα αποτελέσματα για τις γενικές κατηγορίες (εικόνα, όπου περίπου το 37% αυτών είναι κάτω από την κατηγορία «Κυβέρνηση, Πολιτική και Δημόσια Διοίκηση». Η κατηγορία αυτή περιλαμβάνει ότι αφορά τις δραστηριότητες της κεντρικής κυβέρνησης, των τοπικών αυτοδιοικήσεων και των δημόσιων οργανισμών, από τις δαπάνες μέχρι και τα εκλογικά αποτελέσματα. Τα περισσότερα δεδομένα που περιλαμβάνονται στην συγκεκριμένη κατηγορία αφορούν τις δαπάνες των υπουργείων της κεντρικής κυβέρνησης αλλά και των δημοσίων οργανισμών καθώς και της τοπικής αυτοδιοίκησης. Η δημοσιοποίηση των δαπανών του ευρύτερου αλλά και του στενού δημόσιου τομέα είναι από τους κυριότερους άξονες της πολιτικής προς την διαφάνεια και τη υπευθυνότητα του Ηνωμένου Βασιλείου και ο πίνακας 5.9 περιγράφει αναλυτικά τα ευρήματα της έρευνάς μας για τα σύνολα δεδομένων που αφορούν την κατηγορία «Κυβέρνηση, Πολιτική και Δημόσια Διοίκηση». Μεγάλο κομμάτι των δεδομένων αποτελούν και αυτά που περιγράφουν την οργανωσιακή δομή των υπουργείων και των δημοσίων οργανισμών. Η εικόνα 5.17 μας δείχνει οπτικοποιημένα τα αποτελέσματα του πίνακα 5.6.



Εικόνα 5.16 Οπτικοποίηση γενικών κατηγοριών

Τα δεδομένα της δεύτερης στην σειρά κατηγορίας «Υγεία, Ευημερία και Φροντίδα» είναι το 20% των συνόλων δεδομένων. Η κατηγορία αυτή περιλαμβάνει δεδομένα που περιγράφουν ότι αφορά την υγεία και το σύστημα υγείας στο Ηνωμένο Βασίλειο και την κοινωνική πρόνοια σε όλο της το φάσμα. Ενδεικτικά αναφέρουμε κάποιες υποκατηγορίες της που αφορούν στοιχεία για θεραπείες, θνησιμότητα, μητρότητα, προνόμια κοινωνικής πολιτικής, χρόνους αναμονής σε νοσοκομεία κ.α. Η τρίτη κατηγορία, σε αριθμό συνόλων δεδομένων, περιλαμβάνει στοιχεία για «Εκπαίδευση και δεξιότητες». Είναι το 9% του συνολικού αριθμού και τα δεδομένα της έχουν στοιχεία για το εκπαιδευτικό σύστημα, την εκπαίδευση, τα

σχολεία σε όλες τις βαθμίδες τους, τα πανεπιστήμια, επιδόσεις ομάδων μαθητών, κ.ο.κ.

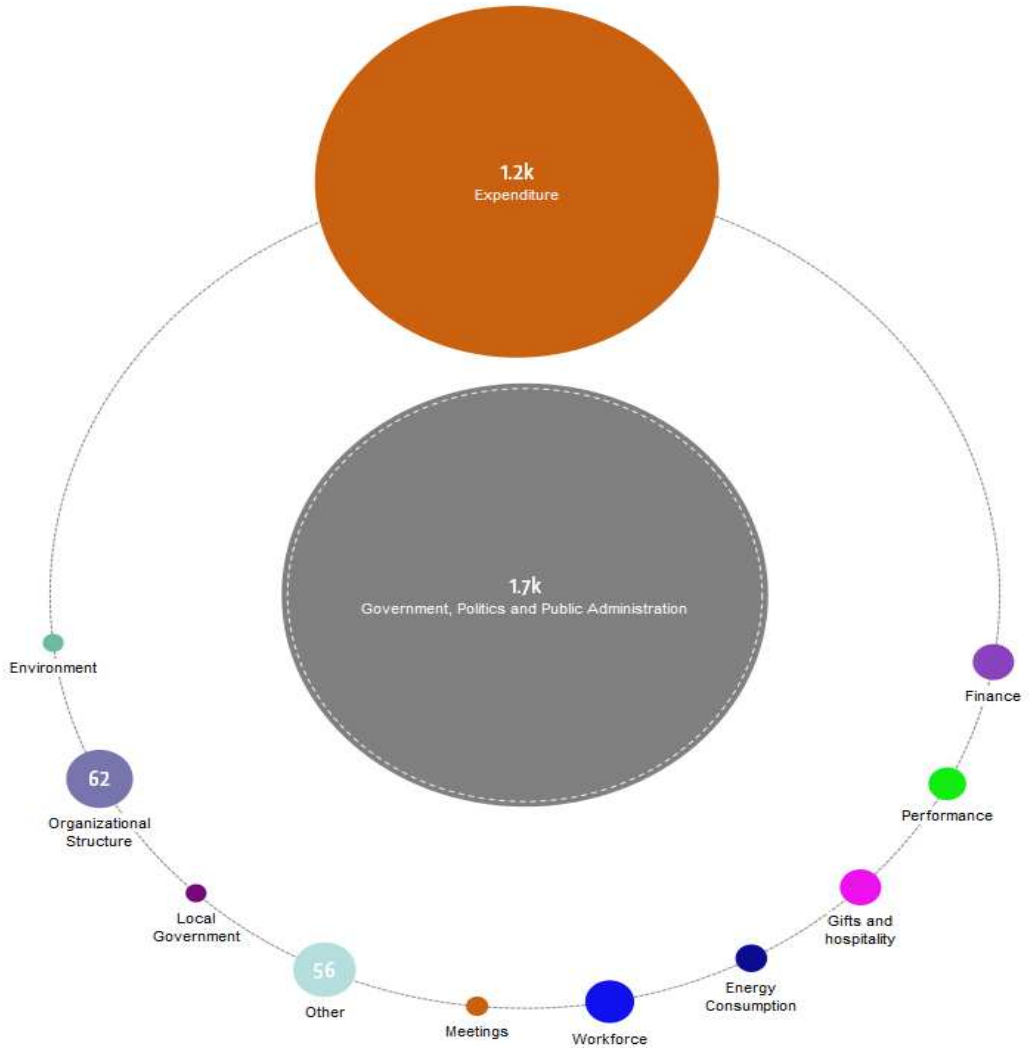
Οι υπόλοιπες κατηγορίες είναι όλες μαζί το 34% των δεδομένων και περιγράφουν διάφορες πτυχές των δραστηριοτήτων του δημόσιου τομέα αλλά και των πολιτών. Περιβαλλοντικά στατιστικά στοιχεία για την διαχείριση των αποβλήτων, την μόλυνση του περιβάλλοντος και παρόμοια δεδομένα περιέχονται στην κατηγορία «Περιβάλλον». Στοιχεία για ποσοστά εγκληματικότητας, για αποφάσεις δικαστηρίων και για κρατούμενους φυλακών περιλαμβάνονται στην κατηγορία «Δημόσια Τάξη, Δικαιοσύνη και Δικαιώματα», επίσης δημογραφικά στοιχεία και δεδομένα για την ζωή στην τοπική κοινωνία είναι κάτω από την κατηγορία «Η ζωή στην κοινότητα».

Government, Politics and Public Administration:	**1696**
Expenditure:	1245
Organizational Structure:	232
Workforce:	37
Finance:	28
Gifts and Hospitality:	28
Performance:	24
Energy Consumption:	18
Meetings:	10
Environment:	9
Local Government:	9
Other:	56

Πίνακας 5.9 Ανάλυση υποκατηγοριών δεδομένων κατηγορίας «Κυβέρνηση, Πολιτική και Δημόσια Διοίκηση»

Ιδιαίτερο ενδιαφέρον έχει να μελετήσουμε τα αποτελέσματα για τις οντότητες που περιγράφουν τα σύνολα δεδομένων. Ως οντότητα ορίζουμε αυτό που υπάρχει ως μια ιδιαίτερη και διακριτή μονάδα, δηλαδή στην περίπτωσή μας αυτό περιγράφεται στην στήλη ενός πίνακα των στατιστικών δεδομένων και μπορεί να χρησιμοποιηθεί ως βάση για τον συνδυασμό διαφορετικών δεδομένων. Στον πίνακα 5.10 μπορούμε να δούμε ποιές οντότητες βρήκαμε μελετώντας τα δεδομένα και στη εικόνα 5.18 την

οπτικοποίηση τους. Στην κορυφή του πίνακα 5.10 βλέπουμε ότι είναι τα κυβερνητικά τμήματα και οι δημόσιοι οργανισμοί με σύνολο 1520 σύνολα δεδομένων.



Εικόνα 5.17 Οπτικοποίηση κατηγορίας «Κυβέρνηση, Πολιτική και Δημόσια Διοίκηση»

Σε αυτήν την κατηγορία των οντοτήτων περιλαμβάνονται τα υπουργεία αλλά και οι δημόσιοι οργανισμοί και η συγκέντρωση τους σε μία κατηγορία έγινε για πρακτικούς λόγους εμφάνισης των αποτελεσμάτων. Το μεγαλύτερο κομμάτι των δεδομένων (1138 στον αριθμό) που περιγράφουν αυτές τις οντότητες αφορούν δαπάνες των κυβερνητικών και δημόσιων οργανισμών και 232 σύνολα δεδομένων αφορούν την οργανωσιακή δομή των υπουργείων και οργανισμών. Μια δεύτερη μεγάλη ομάδα οντοτήτων είναι οι αρχές της τοπικής αυτοδιοίκησης για τις οποίες υπάρχουν 1298 σύνολα δεδομένων που τις περιγράφουν. Στοιχεία από σχεδόν όλες

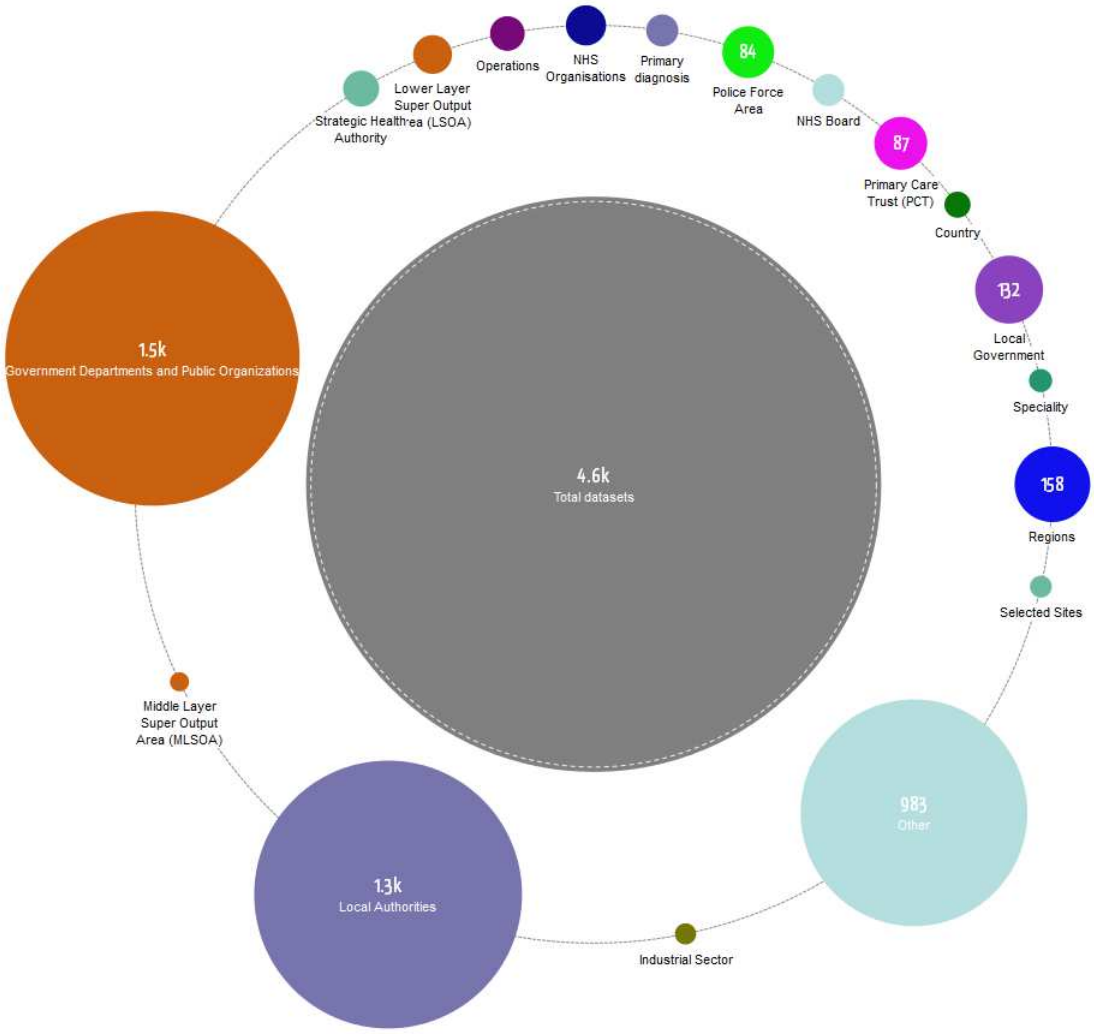
τις βασικές κατηγορίες περιλαμβάνονται στα δεδομένα των τοπικών αρχών με το μεγαλύτερο μέρος τους, περίπου 27%, να αφορά την κατηγορία «Υγεία, ευημερία και φροντίδα» και το ίδιο περίπου ποσοστό αφορά την κατηγορία «Εκπαίδευση και δεξιότητες».

Entities (total number 235):	**4638**
Government Departments and Public Organizations:	1520
Local Authorities:	1298
Regions:	158
Local Government:	132
Primary Care Trust (PCT):	87
Police Force Area:	84
NHS Organisations:	55
Lower Layer Super Output Area (LSOA):	52
Strategic Health Authority:	46
Operations:	43
NHS Board:	38
Primary diagnosis:	38
Country:	27
Speciality:	22
Selected Sites:	20
Industrial Sector:	19
Middle Layer Super Output Area (MLSOA):	16
Other:	983

Πίνακας 5.10 Ανάλυση οντοτήτων συνόλων δεδομένων

Στον πίνακα 5.10 παρατηρούμε ότι υπάρχουν κάποιες οντότητες οι οποίες χαρακτηρίζουν μόνο το Ηνωμένο Βασίλειο. Αυτές είναι κατ` αρχήν οι οργανισμοί Primary Care Trust οι οποίοι ανήκουν διοικητικά στην Αγγλία και είναι 151 σε αριθμό. Αυτοί οι οργανισμοί είναι το κέντρο του συστήματος υγείας, διαχειρίζονται το 80%

του προϋπολογισμού του και επίσης διαχειρίζονται τις υπηρεσίες προς τους ασθενείς που αφορούν τον πρώτο γιατρό που θα επισκεφθεί κάποιος αν έχει κάποιο



Εικόνα 5.18 Οπτικοποίηση οντοτήτων

πρόβλημα υγείας (About the NHS, 2012). Οι οργανισμοί NHS Boards είναι η διοικητική διαίρεση του συστήματος υγείας της Σκωτίας (NHS Scotland, 2012). Τέλος πρέπει να εξηγήσουμε τι είναι οι Super Output Areas, οι οποίες δημιουργήθηκαν αρχικά για να χρησιμοποιηθούν από την ιστοσελίδα των Neighborhood Statistics (NeSS) και είναι γεωγραφικές διαιρέσεις που σχεδιάστηκαν για να βελτιώσουν την αναφορά των στατιστικών στοιχείων μικρών ζωνών. Υπάρχουν τρία επίπεδα Super Output Areas, το πρώτο είναι το Lower Layer όπου το μικρότερο επίπεδο πληθυσμού είναι 1500 κάτοικοι και αποτελείται από ένα γκρουπ από Output Areas, συνήθως πέντε. Το δεύτερο είναι το Middle Layer με μικρότερο επίπεδο πληθυσμού 5000

κατοίκων και αποτελείται από ομάδες Lower Super Output Areas. Το τελευταίο επίπεδο είναι το Upper Layer το οποίο σχεδιάστηκε αρχικά αλλά τελικά δεν υλοποιήθηκε (Super Output Areas, n.d.).

Οι οντότητες ουσιαστικά είναι το κοινό χαρακτηριστικό των συνόλων δεδομένων το οποίο ψάχνουμε για να μπορέσουμε να συνθέσουμε τα δεδομένα. Είναι η βάση πάνω στην οποία μπορούμε να χτίσουμε τις σχέσεις των δεδομένων, όπως είδαμε στο προηγούμενο κεφάλαιο, και να μπορέσουμε να κάνουμε παρατηρήσεις συνθέτοντας σύνολα δεδομένων από ίδιες κατηγορίες (πίνακες 5.4 και 5.5) ή από διαφορετικές κατηγορίες (πίνακες 5.2, 5.3 και 5.6)

5.2.7. ΟΠΤΙΚΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ

Για να καταφέρουμε να κάνουμε τα αποτελέσματα της έρευνας πιο κατανοητά επιλέξαμε να οπτικοποιήσουμε τα δεδομένα, με σκοπό ο τελικός χρήστης να μπορεί να πλοηγείται ευκολότερα στις σχέσεις των συνόλων δεδομένων και τελικά να μπορεί να επιλέξει τις κατηγορίες που θέλει να συνθέσει με απλό τρόπο.

Από τα διάφορα εργαλεία οπτικοποίησης επιλέξαμε το bubble tree visualization το οποίο έχει φτιαχτεί αρχικά από τον David McCandless και επανασχεδιάστηκε από τον Gregor Aisch για το project του Open Knowledge Foundation “Where Does my Money Go”. Οι λόγοι της επιλογής του συγκεκριμένου εργαλείου είναι πρώτον ότι οπτικοποιεί ιεραρχικά δεδομένα, δεύτερον ότι χρησιμοποιείται ήδη από μία εφαρμογή για να οπτικοποιήσει δημόσια ανοιχτά δεδομένα και τρίτον γιατί είναι εφαρμογή ανοιχτού κώδικα. Είναι φτιαγμένο με Javascript , διαθέσιμο ως JavaScript library, χτισμένο πάνω στην Raphael JS και αναπτύσσεται σε html ιστοσελίδα. Το εργαλείο ενσωματώθηκε στο openspending.org με σκοπό να μπορεί κάποιος που θέλει να το χρησιμοποιήσει για να δει τις δαπάνες κάποιας από τις χώρες που υπάρχουν ή και να μπορέσει να ανεβάσει δεδομένα στην ιστοσελίδα (Bubble Tree Library, [n.d.]; Visualize Hierarchical Data with a Bubble Tree, [n.d.]; An open platform for government expenditure tracking, [n.d]).

Η εφαρμογή χρησιμοποιεί φούσκες σαν εικόνες των οποίων το μέγεθος είναι ανάλογο με τον αριθμό των δεδομένων που περιγράφουνε. Υπάρχουν φούσκες δορυφόροι οι οποίοι είναι ιεραρχικά στο παρακάτω επίπεδο και πατώντας πάνω

σ`αυτές που έχουν μία κυκλική διακεκομμένη γραμμή, μπορεί κάποιος να πλοηγηθεί σε πληροφορίες που υπάρχουν στο παρακάτω επίπεδο. Η φούσκα έρχεται στο κέντρο της οθόνης και αριστερά είναι εμφανής η φούσκα του ανώτερου επιπέδου και πάνω και κάτω οι δορυφόροι του ίδιου επιπέδου. Ο κώδικας και τα αρχεία που είναι απαραίτητα για την δημιουργία μιας ανάλογης οπτικοποίησης είναι διαθέσιμα στον σύνδεσμο https://bitbucket.org/okfn/openspendingjs/src/tip/app/bubbletree/ και η τεκμηρίωση στον σύνδεσμο http://vis4.net/blog/en/posts/tutorial-bubble-tree/ .

Για την υλοποίηση της οπτικοποίησης χρειάστηκε να παραμετροποιήσουμε τα δεδομένα μας για να μπορέσουν να εμφανιστούν με τον σωστό τρόπο από το εργαλείο. Για την εισαγωγή δεδομένων υπάρχουν τρείς τρόποι α) κατασκευή ενός .json αρχείου μέσα στο οποίο πρέπει να υπάρχουν τα δεδομένα κατάλληλα δομημένα σύμφωνα με την ταξινόμηση που δίνεται από το αρχείο cofog.js, το οποίο αρχείο βρίσκεται τοπικά στον υπολογιστή β) εισαγωγή δεδομένων χειροκίνητα στην var data στο js script που βρίσκεται στο αρχείο index, τα οποία πρέπει επίσης να είναι ταξινομημένα κατάλληλα σύμφωνα με την ταξινόμηση του cofog.js, γ) εισάγοντας έναν σύνδεσμο για κάποια διεύθυνση στο διαδίκτυο όπου υπάρχει ένα σύνολο δεδομένων κατάλληλα φτιαγμένο για οπτικοποίηση. Η δικιά μας επιλογή ήταν η χειροκίνητη εισαγωγή δεδομένων και κάναμε και μία μικρή διαμόρφωση των χρωμάτων που εμφανίζονται στις φούσκες. Παραδείγματα της οπτικοποίησης βλέπουμε στις εικόνες 5.9 έως 5.14 και τον τρόπο ταξινόμησης των δεδομένων τον εξηγήσαμε σε προηγούμενο κεφάλαιο.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

Οι ανάγκες των πολιτών οδήγησαν τις σημερινές κυβερνήσεις σε νέα μοντέλα διακυβέρνησης, τα οποία λειτουργούν με μεγαλύτερη διαφάνεια και υπευθυνότητα. Αρωγός στα νέα αυτά μοντέλα είναι η αλματώδης εξέλιξη στο τομέα των τηλεπικοινωνιών και των ηλεκτρονικών υπολογιστών. Κομμάτι των νέων ειδών διακυβέρνησης είναι η δημοσίευση των δεδομένων που συγκεντρώνει ο δημόσιος τομέας, σε ελεύθερη μορφή για όποιον θέλει να τα χρησιμοποιήσει ή να τα μελετήσει.

Στην βιβλιογραφική επισκόπηση που προηγήθηκε εξετάσαμε και αναλύσαμε την έννοια της ανοιχτής διακυβέρνησης και τις έννοιες και τους ορισμούς των δημόσιων ανοιχτών δεδομένων με σκοπό να κατανοηθεί από τον αναγνώστη η σχέση τους και ο ρόλος που παίζουν τα δημόσια ανοιχτά δεδομένα στην διαφάνεια και την υπευθυνότητα των κυβερνήσεων απέναντι στους πολίτες.

Εξετάζοντας τις βασικές αρχές των δημόσιων ανοιχτών δεδομένων παρατηρήσαμε πότε τεχνικά τα δεδομένα θεωρούνται ότι είναι ελεύθερα και ανοιχτά στον πιθανό χρήστη και κάτω από ποιούς κανόνες πρέπει οι κυβερνήσεις να δημοσιοποιούν τα δεδομένα τους. Επίσης μελετήσαμε τις πρωτοβουλίες των κυβερνήσεων ανά τον κόσμο για την δημοσιοποίηση των δεδομένων τους μέσω καταλόγων και την εμπορική αξία που μπορούν να πάρουν αυτά τα δεδομένα εάν επαναχρησιμοποιηθούν και χτιστούν πάνω σ`αυτά εφαρμογές οι οποίες θα εξυπηρετούν τους πολίτες. Η Ευρωπαϊκή Ένωση έδειξε τον δρόμο, με τις ντιρεκτίβες που έχει εκδώσει, στις ευρωπαϊκές κυβερνήσεις πως να δημοσιοποιήσουν ελεύθερα τα δεδομένα τους.

Η εμπορική αξία μέσω της επαναχρησιμοποίησης καθοδήγησε και την έρευνα μας για την χαρτογράφηση των δημοσίων δεδομένων του Ηνωμένου Βασιλείου. Η ιστοσελίδα http://www.data.gov.uk βρέθηκε από την βιβλιογραφική επισκόπηση να είναι η πιο συμβατή με τις αρχές και τους νόμους των δημόσιων ανοιχτών δεδομένων και η Βρετανική κυβέρνηση σε μεγάλο βαθμό αφοσιωμένη στην ανοιχτή διακυβέρνηση. Ο σκοπός της έρευνας ήταν να μελετηθούν οι σχέσεις μεταξύ των συνόλων δεδομένων, να καταγραφούν και να γίνει δυνατό να επαναχρησιμοποιηθούν συνθέτοντας τα μεταξύ τους. Μελετήσαμε όλα τα σύνολα

δεδομένων, τα κατηγοριοποιήσαμε και παρουσιάσαμε μία οπτικοποιημένη ανάλυση σαν οδηγό για τον μελλοντικό χρήστη που επιθυμεί να μελετήσει ή να χτίσει εφαρμογές πάνω σ`αυτά και να τους προσθέσει εμπορική αξία. Τα σύνολα δεδομένων από μόνα τους μπορούν να προσφέρουν χρήσιμες πληροφορίες, η σύνθεσή τους όμως με βάση κάποια κοινή οντότητα μπορεί να οδηγήσει σε πολύτιμα συμπεράσματα για την λήψη αποφάσεων. Μέσω της έρευνας καταγράψαμε αυτές τις σχέσεις σαν έναν χάρτη για τον πιθανό ενδιαφερόμενο. Επίσης ενδιαφέρον έχει και το περιεχόμενο των δεδομένων ως μελέτη για αυτούς που ενδιαφέρονται να δουν πως ο πιο συμβατός κατάλογος με τις αρχές των ανοιχτών δεδομένων λειτουργεί και τι πληροφορίες δημοσιεύει.

Σύμφωνα με την έρευνα τα συμπεράσματα που μπορέσαμε να εξάγουμε είναι τα παρακάτω:

- Ο κατάλογος των ανοιχτών δημόσιων δεδομένων του Ηνωμένου Βασιλείου είναι συμβατός με τις αρχές των ανοιχτών δεδομένων, καθώς μόνο το 6% των δεδομένων του δεν είναι αναγνώσιμα από υπολογιστή.
- Είναι αρκετά λειτουργικός και ενημερωμένος.
- Έχει ποικιλία στατιστικών δεδομένων, 61% μαζί με τις δαπάνες του δημοσίου τομέα, που έχουν δημοσιευθεί από τα περισσότερα τμήματα της κυβέρνησης, υπουργεία και οργανισμούς. Το 12% είναι γεωχωρικά δεδομένα.
- Έχει όμως και δεδομένα που δεν μπορεί κάποιος να έχει πρόσβαση και αυτά αποτελούν το 17% του συνόλου των δεδομένων. Επίσης το 5% των συνόλων δεδομένων οδηγούν σε ιστοσελίδες όπου βγαίνει το μήνυμα 404 not found, το 2% είναι περιορισμένης πρόσβασης και το 2% είναι κενά. Άρα παρόλο που είναι ο πιο συμβατός κατάλογος με τα ανοιχτά δεδομένα υπάρχουν προβλήματα προς επίλυση.
- Σύμφωνα με την πολιτική της κυβέρνησης του Ηνωμένου Βασιλείου για την διαφάνεια, τα περισσότερα σύνολα δεδομένων αφορούν τις δαπάνες των υπουργείων και των δημόσιων οργανισμών και είναι το 26% των στατιστικών δεδομένων.
- Η επαναχρησιμοποίηση και σύνθεση των δεδομένων μέσα στην πληθώρα που υπάρχει δεν είναι εύκολη καθώς τα μεταδεδομένα δεν είναι

ολοκληρωμένα και χρειάζεται αρκετό ψάξιμο και δουλειά για να γίνει αυτό εφικτό.

Το γεγονός ότι το ρεύμα προς τα ανοιχτά δημόσια δεδομένα εντάθηκε την τελευταία πενταετία, παρά την διάθεση των κυβερνήσεων για δημοσιοποίηση των δεδομένων τους, βρήκε την δομή του δημόσιου τομέα σχετικά ανέτοιμη. Η γραφειοκρατία και οι παλιές νοοτροπίες των δημοσίων υπαλλήλων βάζουν εμπόδια στην δημοσιοποίηση των κατάλληλων αρχείων ή δεδομένων και επειδή οι αλλαγές σε οργανισμούς του δημοσίου δεν γίνονται γρήγορα, υπάρχουν προβλήματα και καθυστερήσεις. Ο όγκος των πληροφοριών που συλλέγονται σε πολλές περιπτώσεις θεωρείται περιουσιακό στοιχείο του οργανισμού ή του φορέα που τα συλλέγει και υπάρχει αντίσταση στην λογική να διαθέτονται ελεύθερα στο διαδίκτυο σε αντίθεση με την λογική του Tim Berners-Lee ο οποίος υποστηρίζει ότι για την συλλογή των κυβερνητικών δεδομένων έχουν ήδη δαπανηθεί χρήματα και αν κάθονται στο δίσκο ενός υπολογιστή πάνε χαμένα.

Βλέπουμε λοιπόν ότι από την μία πλευρά έχουμε τους υποστηρικτές του «ανοίγματος» των δημόσιων ή κυβερνητικών δεδομένων, που είναι διάφορες κυβερνητικές ή ιδιωτικές πρωτοβουλίες, και από την άλλη έχουμε την αντίσταση των γραφειοκρατικά δομημένων δημόσιων οργανισμών και των παλαιών νοοτροπιών. Οι πρώτοι υποστηρίζουν ότι μέσα από την ελευθερία της γνώσης και της πληροφορίας μπορούμε να οδηγηθούμε σε καλύτερα μοντέλα διακυβέρνησης όπου θα υπάρχει διαφάνεια, υπευθυνότητα, μεγαλύτερη συμμετοχή από τους πολίτες στα κοινά και θα προστεθεί εμπορική αξία σε δεδομένα που κάθονται ανεκμετάλλευτα. Στους δεύτερους συναντάμε την κλασσική αντίσταση στην αλλαγή και στην διάθεση να μοιραστούν κάτι που νομίζουν ότι τους ανήκει. Ο δρόμος για την επίτευξη της σύγκλισης των δύο μερών είναι το νομοθετικό πλαίσιο το οποίο και θα πιέσει προς την μεριά της δημοσιοποίησης αλλά και θα προτυποποιήσει τον τεχνικό τρόπο της δημοσιοποίησης. Η έλλειψη της σωστής καθοδήγησης απλά μπερδεύει ακόμη και αυτούς που έχουν την διάθεση να εμπλακούν και βοηθάει αυτούς που αντιστέκονται.

Το κενό που υπάρχει στο νομοθετικό πλαίσιο, προσπαθεί να το καλύψει η Ευρωπαϊκή Ένωση δίνοντας κατευθύνσεις και προσπαθώντας να δεσμεύσει τις

κυβερνήσεις με χρονοδιαγράμματα υλοποίησης που καταγράφονται στις ντιρεκτίβες. Είναι απαραίτητη όμως και η διάθεση των κυβερνήσεων, η οποία φαίνεται ότι αυξάνεται τα τελευταία χρόνια, όμως χρειάζεται αρκετός δρόμος και προσπάθεια προς την κατεύθυνση ενός ολοκληρωμένου σχεδίου που θα καλύπτει όλες τις πλευρές. Μία πολύ σοβαρή πτυχή που θα πρέπει να ληφθεί σοβαρά υπόψη είναι τα ευαίσθητα δεδομένα και πως αυτά θα διαφυλαχθούν από το άνοιγμα των δεδομένων. Το θέμα πρέπει να εξετασθεί με μεγάλη προσοχή και να προστατευθούν πληροφορίες οι οποίες είτε είναι πολύτιμες για το κράτος είτε είναι προσωπικά δεδομένα αλλά και να μην εμποδιστεί η ελεύθερη διάδοση της γνώσης. Όσο αντιφατικό και αν ακούγεται αυτό που μόλις περιγράψαμε, είναι με λίγα λόγια το πλαίσιο μέσα στο οποίο πρέπει να κινηθούν οι νομικοί που θα ασχοληθούν με το θέμα.

Όσον αφορά το τεχνικό κομμάτι της δημοσιοποίησης των δεδομένων, εκεί οι γραμμές είναι πιο ξεκάθαρες και ίσως είναι και ο τομέας με τον οποίο ασχολήθηκαν περισσότερο οι υπεύθυνοι των κυβερνήσεων. Άρα οι τρόποι δημοσίευσης των δεδομένων είναι πιο ξεκάθαροι και πιο εύκολο να ακολουθηθούν και όπως βλέπουμε στην έρευνα που κάναμε στον Βρετανικό κατάλογο δημοσίευσης δεδομένων, ένα πολύ μεγάλο ποσοστό των συνόλων δεδομένων είναι αναγνώσιμο από υπολογιστή και κάτω από ανοιχτά πρότυπα ακολουθώντας τις βασικές αρχές για να θεωρείται ένα σύνολο δεδομένων ανοιχτό. Τα τεχνικά θέματα που χρειάζονται βελτίωση είναι η ποιότητα της πληροφορίας και οι πόροι για να είναι ικανοί οι δημόσιοι οργανισμοί να δημοσιεύουν τα δεδομένα τους.

Μελλοντικά υπάρχουν πολλά θέματα με τα οποία θα μπορούσε να ασχοληθεί κάποιος. Το γεγονός ότι το θέμα των ανοιχτών δημόσιων δεδομένων είναι σχετικά καινούργιο και αρκετά ενδιαφέρον, σημαίνει ότι η έρευνα για αυτό εξελίσσεται πολύ γρήγορα. Οι αρχές των ανοιχτών δεδομένων εμπλουτίζονται και καθορίζουν πιο συγκεκριμένα τους τρόπους δημοσίευσής τους. Η περαιτέρω διερεύνηση των τρόπων που τα δεδομένα πρέπει να δημοσιεύονται και την ποιότητα της πληροφορίας που πρέπει να περιέχουν θα διαλευκάνει ακόμη περισσότερο το τοπίο και θα βοηθήσει τους εμπλεκόμενους να ακολουθήσουν τον δρόμο των ανοιχτών δεδομένων. Η εμπορική αξία των δεδομένων είναι ένας σημαντικός τομέας διερεύνησης, καθώς ο αριθμός των πρωτοβουλιών μεγαλώνει

ανά τον κόσμο, κατά συνέπεια και ο αριθμός των δεδομένων που δημοσιοποιούνται, τα νούμερα αλλάζουν και η σύνδεση αυτής της προστιθέμενης αξίας στα δεδομένα με τις εφαρμογές που έχουν φτιαχτεί πάνω σ` αυτά και τους πιθανούς τρόπους χρήσης τους θα ήταν χρήσιμα. Η χαρτογράφηση των ελληνικών δημόσιων δεδομένων και το επίπεδο δέσμευσης της ελληνικής κυβέρνησης με τις αρχές της ανοιχτής διακυβέρνησης θα μπορούσε να είναι αντικείμενο μελλοντικής μελέτης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

8 Principles of Open Government Data [online]. (2007) [Accessed 30/9/2011]. Available from: http://www.opengovdata.org/home/8principles

About the NHS [online]. (2012) [Accessed 31/7/2012]. Available from: http://www.nhs.uk/NHSEngland/thenhs/about/Pages/authoritiesandtrusts.aspx#primary

Access to information: a fundamental right, a universal standard [online]. (2006) [Accessed 15/12/2011]. Available from: http://www.accessinfo.org/documents/Access_Docs/Thinking/Get_Connected/Access_Info_Europe_Briefing_Paper.pdf

Alonso, J. et al. (2009). *Improving Access to Government through Better Use of the Web* [online]. [Accessed 31/8/2011]. Available from: http://www.w3.org/TR/2009/NOTE-egov-improving-20090512/

An open platform for government expenditure tracking [online]. (n.d.) [Accessed 31/7/2012]. Available from: http://openspending.org/help/index.html

Annotations to the 8 Principles [online]. (2007) [Accessed 30/9/2011]. Available from: http://www.opengovdata.org/home/8principles/annotations

Audit Commission (2009). *Nothing but the Truth? A Discussion Paper* [online]. [Accessed 15/12/2011]. Available from: http://www.audit-commission.gov.uk/nationalstudies/localgov/Pages/nothingbutthetruth.aspx

Bennett, D. and Harvey, A. (2009). *Publishing Open Government Data* [online]. [Accessed 30/9/2011]. Available from: http://www.w3.org/TR/gov-data/

Berners-Lee, T. (2006). *Linked Data* [online]. [Accessed 30/9/2011]. Available from: http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T. (2009). *Putting Government Data online* [online]. [Accessed 31/8/2011]. Available from: http://www.w3.org/DesignIssues/GovData.html

Berners-Lee, T. (2010). *Design Issues: Linked Data* [online]. [Accessed 31/8/2011]. Available from: http://www.w3.org/DesignIssues/LinkedData.html

Bizer, C., Heath, T. and Berners-Lee, T. (2009). Linked data – the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*. 5, pp.1–22, Special Issue on Linked Data.

Bubble Tree Library [online]. (n.d.) [Accessed 31/7/2012]. Available from: http://driven-by-data.net/about/interactive-bubbletree/#/1

Cabinet Office (2012). *UK Government Reference Architecture (UKRA)* [online]. [Accessed 30/5/2012]. Available from: http://www.cabinetoffice.gov.uk/sites/default/files/resources/UK-Reference-Architecture-V1-0-HMG-Branded.pdf

Central Office of Information (COI) (2011). *Underlying Data Publication* [online]. [Accessed 15/12/2011]. Available from: http://coi.gov.uk/guidance.php?page=376

Chen, J. (2009). Qualitative Evaluation of E-Government Service based on Participatory Index and Balanced Scorecard Approach. *In: International Conference on Information Management, Innovation Management and Industrial Engineering*, 26-27 Dec., Xi'an. pp.380 - 383.

Cornell University (n.d.). *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems* [online]. [Accessed 15/12/2011]. Available from: http://www.dpworkshop.org/dpm-eng/eng_index.html

Cyganiac, R. (2011 November). *How to publish Open Data*. Presentation presented at Deri University, Galway, Ireland.

Data.gov.uk [online]. (2012) [Accessed 31/7/2012]. Available from: http://en.wikipedia.org/wiki/Data.gov.uk

Davenport, T. and Prusak, L. (1998). *Working Knowledge: How Organizations Manage What They Know.* Boston: Harvard Business Press, p.2 [Accessed 15/12/2011]. Available from: http://books.google.com/books?id=-4-7vmCVG5cC&printsec=frontcover&dq=isbn:9780875846552&hl=el&ei=y5_gTv2xLMjYsgafrdD8CA&sa=X&oi=book_result&ct=book-preview-link&resnum=1&ved=0CDIQuwUwAA#v=onepage&q&f=false

Davies, T. (2010a). *Open Data, Democracy and Public Sector Reform. A Look at Open Government Data Use from data.gov.uk, based on MSc Dissertation* [online]. [Accessed 31/10/2011]. Available from: http://practicalparticipation.co.uk/odi/report/

Davies, T. (2010b). *The current push to make public sector information available as raw data in open digital online forms will radically transform research. Discuss, with reference to a number of case studies*. Essay for MSc Social Science of the Internet thesis, Oxford Internet Institute.

Dias, G.P. and Rafael, J.A. (2007). A simple model and a distributed architecture for realizing one-stop e-government. *Electronic Commerce Research and Applications*. 6, pp.81-90.

Ding, L., et al. (2011). TWC LOGD: A portal for linked open government data ecosystems. *Web Semantics: Science, Services and Agents on the World Wide Web*. 9, pp.325-333.

Eaves, D. (2009). *Three Laws of Open Data (International Edition)* [online]. [Accessed 31/11/2011]. Available from: http://eaves.ca/2009/11/29/three-laws-of-open-data-internationaledition/

Erickson, J., Shi, Y., Ding, L., Rozell, E., Zheng, J. and Hendler, J. (2011). TWC International Open Government Dataset Catalog. *In: Proceedings of the 7th International Conference on Semantic Systems*, 7-9 Sept., Graz, Austria. pp.227-229.

EURIM (2011). *Improving the Evidence Base The Quality of Information.* Status Report and Recommendations of The Information Society Alliance EURIM Sub-group on the Quality of Information.

European Commission (2003). Directive 2003/98/EC of the European parliament and of the council on the re-use of public sector information. Official *Journal of the European Union*. 345, pp.90-96.

European Commission (2009). Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Re-use of Public Sector Information – Review of Directive 2003/98/EC, COM (2009) 212 final, European Commission, Brussels.

European Commission (2010). *A Digital Agenda for Europe. 52010DC0245(01) COM/2010/0245 f/2. Brussels: European Commission* [online]. [Accessed 15/12/2011]. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:REV1:EN:HTML

European Commission (2011). *Digital Agenda: Turning government data into gold* . Press release, issued 12/12/2011 [online]. [Accessed 12/12/2011]. Available from: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/1524&format=HTML&aged=0&language=EN&guiLanguage=en

Geiger, C. and von Lucke, J. (2011). Open Government Data: Free accessible data of the public sector. *In: Proceedings of the International Conference for E-Democracy and Open Government*, 5-6 May, Krems, Danube University Krems, Austria. pp.183-194.

Government launches one-stop shop for data [online]. (2010) [Accessed 31/7/2012]. Available from: http://webarchive.nationalarchives.gov.uk/+/http://www.cabinetoffice.gov.uk/newsroom/news_releases/2010/100121-data.aspx

Gray, J. (2011 April). "*Open Government Data: What, Why, How?*". Presentation presented at Sala delle Colonne, Camera dei Deputati, Rome, Italy.

Halonen, A. (2012). *Being Open About Data: Analysisof the UK open data policies and applicability of open data* [online]. [Accessed 31/5/2012]. Available from: http://www.finnish-institute.org.uk/images/stories/pdf2012/being%20open%20about%20data.pdf

HM Government (2009). *Putting the frontline first: smarter government,* [online]. [Accessed 30/9/2011]. Available from: http://www.hmg.gov.uk/media/52788/smarter-government-final.pdf

Hyland, B., Terrazas, B. and Capadisli, S. (2011). *Linked Data Cookbook* [online]. [Accessed 15/12/2011]. Available from: http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook

Kalampokis, E., Tambouris, E. and Tarabanis, K. (2011a). A classification scheme for open government data: towards linking decentralised data. *Int. J. Web Engineering and Technology*. 6, pp.266-285.

Kalampokis, E., Tambouris, E. and Tarabanis, K. (2011b). Open Government Data: A Stage Model. *Electronic Government, Lecture Notes in Computer Science*. 6846, pp.235-246.

Katleen, J. (2011). The role of public sector information in the European market for online content: a never-ending story or a new beginning?. *Info*. 13, pp.20-29.

Koumenides, C., Salvadores, M., Alani, H. and Shadbolt (2010). Global Integration of Public Sector Information. *In: Web Science Conference*, 26-27 April, Raleigh, NC, USA.

Layne, K. and Lee, J. (2001). Developing fully functional e-government: a four stage model. *Government Information Quarterly*. 18, pp. 122-136.

Lee, J. (2010). 10 year retrospect on stage models of e-government: a qualitative meta-synthesis. *Government Information Quarterly*. 27, pp. 220-230.

Linders, D. and Wilson, S. (2011). What is Open Government? One Year after the Directive. *In: Proceedings of the 12th Annual International Conference on Digital Government Research*, 12-15 June, College Park, MD, USA. pp. 262-271.

Maali, F., Cyganiak, R. and Peristeras, V. (2010). Enabling Interoperability of Government Data Catalogues. *Lecture Notes in Computer Science*. 6228, pp. 339-350.

McDermott, P. (2010). Building Open Government. *Government Information Quarterly*. 27, pp.401-413.

NHS Scotland [online]. (2012) [Accessed 31/7/2012]. Available from: http://en.wikipedia.org/wiki/NHS_Scotland

Obama, B. (2009). *Memorandum for the Heads of executive Departments and Agencies.Transparency and Open Government* [online]. [Accessed 30/9/2011]. Available from: http://www.whitehouse.gov/the_press_office/Transparency_and_Open_Government/

Office of Management and Budget (2009). *Open Government Directive* [online]. [Accessed 15/12/2011]. Available from: http://www.whitehouse.gov/open/documents/open-government-directive

Open Government Data [online]. (2011) [Accessed 30/9/2011]. Available from: http://opengovernmentdata.org/what/

Open Government Data on a Map [online]. (2011) [Accessed 30/9/2011]. Available from: http://opengovernmentdata.org/data/map/

Open Definition [online]. (2011a) [Accessed 30/9/2011]. Available from: http://opendefinition.org/

Open Definition [online]. (2011b) [Accessed 30/9/2011]. Available from: http://opendefinition.org/okd/

O`Reilly, T. (2010). Government As a Platform. *In*: Lathrop, D. and Ruma, D., (ed). *Collaboration, Transperancy, and Participation in Practice*, USA: O`Reilly Media Inc., pp11-39.

Definition of data set [online]. (2012) [Accessed 31/7/2012]. Available from: http://oxforddictionaries.com/definition/english/data%2Bset

Papadopoulos, T. (2011). *Open Public Data*. M.Sc thesis, National School of Public Administration.

Parycek, P. (2010). *Open Government: Information Flow in Web 2.0* [online]. [Accessed 30/9/2011]. Available from: http://www.epractice.eu/en/document/313345

Peled, A. (2011). When Transparency and Collaboration Collide: The USA Open Data Program. *Journal of the American Society for Information Science and Technology*. 62, pp.2085-2094.

Pizzicannella, R. (2010). Co-production and open data: the right mix for public service effectiveness?. *In: draft papers, 10th European Conference on eGovernment*, 17-18 June, Limerick, Ireland.

PIRA (2000), *Commercial exploitation of Europe's public sector information*, *Executive summary*, Pira International Ltd, University of East Anglia and KnowledgeView Ltd, and *Final Report*, Pira International, European Commission, Directorate General for the Information Society.

Pollock, R. (2011). *Welfare gains from opening up Public Sector Information in the UK* [online]. [Accessed 15/12/2011]. Available from: http://rufuspollock.org/economics/papers/psi_openness_gains.pdf

Pollock, R., Newbery, D. and Bently, L. (2008). *Models of public sector information provision via trading funds* [online]. [Accessed 30/9/2011]. Available from: http://www.berr.gov.uk/files/file45136.pdf

Renear A., Sacchi S. and Wickett K. (2010). Definitions of Dataset in the Scientific and Technical Literature. In: ASIST, October 22-27, Pittsburg, PA, USA. Pp.1-4.

Robinson, D., Yu, H., Zeller, W.P. and Feltern, E.W. (2009). Government data and the invisible hand. *Yale Journal of Law and Technology*. 11, pp.160-175.

Rogers, S. (2012). UK Open government data: the results of the official audit. *The Guardian* [online] 18 April. [Accessed 31/7/2012] Available from: http://www.guardian.co.uk/news/datablog/2012/apr/18/uk-open-government-data-national-audit-office

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., and Hall, W. (2012). Linked Open Government Data: Lessons from Data.gov.uk. *IEEE INTELLIGENT SYSTEMS*. **12**, pp.1-10.

Schellong, A. and Stepanets, E. (2011). *Unchartered Waters: The State of Open Data in Europe*. Wiesbaden, Germany: CSC [online]. [Accessed 30/9/2011]. Available from: http://assets1.csc.com/de/downloads/CSC_policy_paper_series_01_2011_unchartered_waters_state_of_open_data_europe_English_2.pdf

Super Output Areas [online]. (n.d.) [Accessed 31/7/2012]. Available from: http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html

Tambouris, E. (2001). An integrated platform for realizing online one-stop government: the eGOVproject. *In: Proceedings of the 12th International Workshop on Database and Expert Systems Applications (DEXA 2001)*, 3-7 September, Munich, Germany. pp.359-363.

Tapscott, D. (2010). Foreword. *In*: Lathrop, D. and Ruma, D., (ed). *Collaboration, Transperancy, and Participation in Practice*, USA: O`Reilly Media Inc., pp.xv-xvii.

Tauberer, J. (2009). *Open Data is Civic Capital: Best Practices for "Open Government Data* [online]. [Accessed 23/12/2011]. Available from: http://razor.occams.info/pubdocs/opendataciviccapital.html

Tim Berners-Lee on the UK national data website launch [online]. (2010) [Accessed 31/7/2012]. Available from: http://www.guardian.co.uk/politics/video/2010/jan/21/uk-national-data-website-launched

Tim Berners-Lee unveils government data project [online]. (2010) [Accessed 31/7/2012]. Available from: http://news.bbc.co.uk/2/hi/technology/8470797.stm

The British Government [online].(2011) [Accessed 15/4/2012]. Available from: http://projectbritain.com/government/government.htm

The public sector transparency board-who`s who [online]. (n.d.) [Accessed 30/11/2011]. Available from: http://www.cabinetoffice.gov.uk/content/public-sector-transparency-board-who%27s-who

Transperancy-Overview [online]. (n.d.) [Accessed 30/11/2011]. Available from: http://www.cabinetoffice.gov.uk/content/transparency-overview

United Nations and American Society for Public Administration (ASPA). (2002). Benchmarking e-government: A global perspective. New York, NY: U.N. Publications.

Veljković, N., Bodganović-Dinić, S. and Stoimenov, L. (2011). Municipal Open Data Catalogues. *In: Proceedings of the International Conference for E-Democracy and Open Government*, 5-6 May, Krems, Danube University Krems, Austria. pp.195-207.

Vickery, G. (2011). *Review of Recent Studies on PSI Re-Use and Related Market Developments* [online]. [Accessed 30/11/2011]. Available from: http://old.epsiplatform.eu/news/news/review_of_recent_psi_re_use_studies_published

Visualize Hierarchical Data with a Bubble Tree [online]. (n.d.) [Accessed 31/7/2012]. Available from: http://thewhyaxis.info/bubble-tree/

Weinberger, D. (2009). Objectivity, Transparency, and the Engagement with Democracy. *In*: Gøtze J. and Pedersen, C., (ed). *State of the eUnion*, Copenhagen: 21gov.net., pp. 189-195 [Accessed 31/10/2011]. Available from: http://21gov.net/wp-content/uploads/e-book.pdf

Where does my money go? [online]. (2011) [Accessed 31/12/2011]. Available from: http://wheredoesmymoneygo.org/about/

White House (2011). *The Obama Administration`s Commitment to Open Government: A Status Report,* [online]. [Accessed 17/9/2011]. Available from: http://www.whitehouse.gov/sites/default/files/opengov_report.pdf

Yildiz, M. (2007). E-government research: Reviewing the literature, limitations, and ways forward. *Government Information Quarterly* . 24, pp. 646-665.

Yin, R. (2009). *Case Study Research Design and Methods.* USA: SAGE Publications, Inc. [Accessed 31/3/2012]. Available from: http://www.amazon.com/Case-Study-Research-Methods-Applied/dp/1412960991/ref=pd_rhf_pe_s_cp_1#reader_1412960991 .

.