



Τμήμα Μηχανικών Πληροφοριακών &
Επικοινωνιακών Συστημάτων

Οπτικοποίηση Δεδομένων Ηλεκτρονικής Διακυβέρνησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

των

Μπακόλα Ιωάννη

Τουρνόγλου Δημήτριου

Επιβλέπων: Χαραλαμπίδης Ιωάννης

Επ.Καθηγητής

Τμήματος Μ.Π.Ε.Σ

Σάμος, Σεπτέμβριος 2013

Η ΔΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΔΙΔΑΣΚΟΝΤΩΝ ΕΓΚΡΙΝΕΙ

ΤΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

των

ΜΠΑΚΟΛΑ ΙΩΑΝΝΗ

ΤΟΥΡΝΟΓΛΟΥ ΔΗΜΗΤΡΗ

Σάμος, Σεπτέμβριος 2013

ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ

Δρ ΙΩΑΝΝΗΣ ΧΑΡΑΛΑΜΠΙΔΗΣ, Επιβλέπων

Τμήμα Μηχανικών Πληροφοριακών και

Επικοινωνιακών Συστημάτων

Δρ ΚΟΚΟΛΑΚΗΣ ΣΠΥΡΟΣ, Μέλος

Τμήμα Μηχανικών Πληροφοριακών και

Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΕΥΧΑΡΙΣΤΙΕΣ

Θα θέλαμε να ευχαριστήσουμε τον καθηγητή μας Δρ. Ιωάννη Χαραλαμπίδη (Επίκουρο Καθηγητή του Τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων) για την καθοδήγηση, τις πολύτιμες συμβουλές και επισημάνσεις που μας προσέφερε κατά τη διάρκεια της εκπόνησης της διπλωματικής μας εργασίας.

Ευχαριστούμε θερμά το κ. Χαράλαμπο Αλεξόπουλο και την κ. Ανδρουτσοπούλου Αγγελική για την επικοινωνιακή συνεργασία αλλά και την πολύτιμη βοήθεια που μας παρείχαν στην επίλυση των προβλημάτων που αντιμετωπίσαμε.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

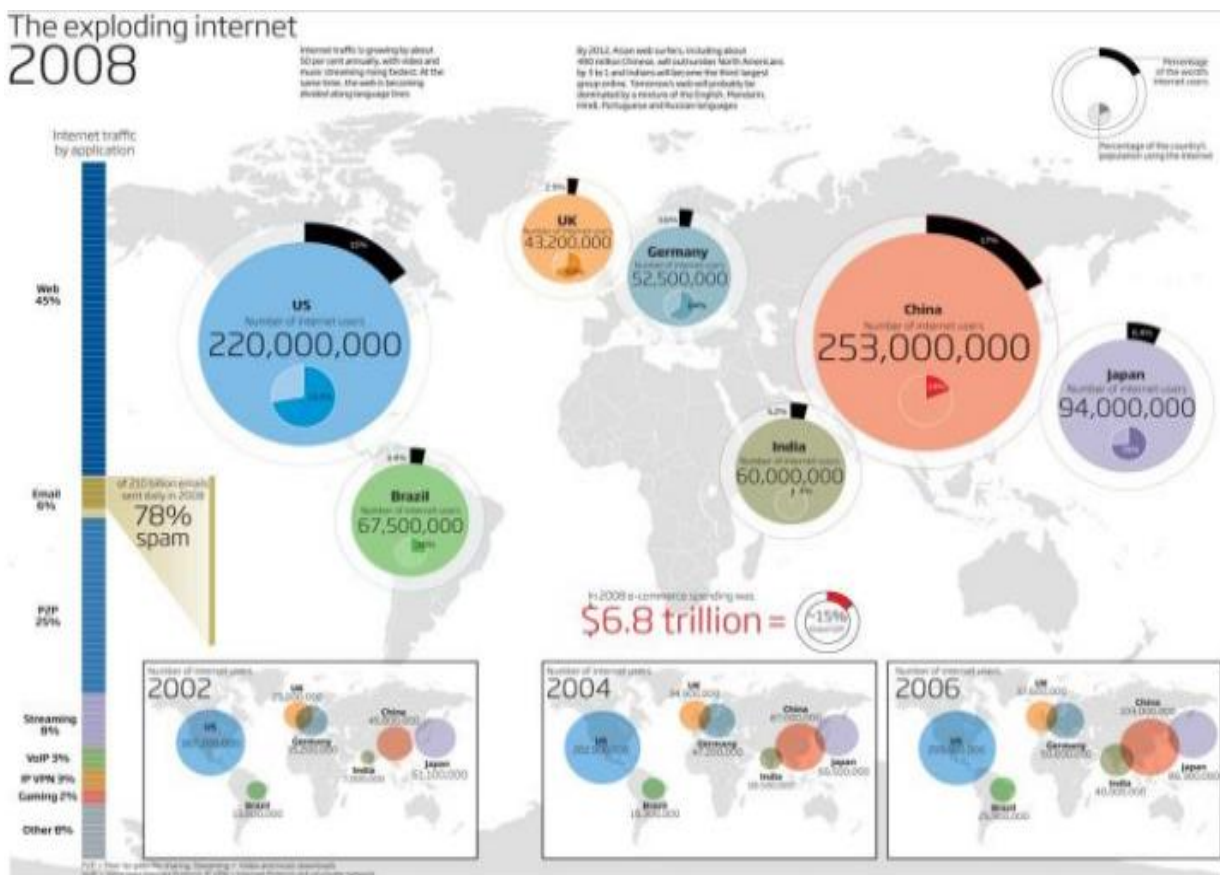
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ.....	5
ΚΕΦΑΛΑΙΟ 1 - Εισαγωγή.....	7
ΚΕΦΑΛΑΙΟ 2 - Ανοικτά Δημόσια Δεδομένα.....	10
2.1 Ορισμοί.....	11
2.1.1 Δεδομένα και Πληροφορίες.....	11
2.1.2 Το Θεμελιώδες τετράπτυχο «Δεδομένα -Πληροφορία-Γνώση – Σοφία».....	12
2.1.3 Ανοικτά Δημόσια Δεδομένα.....	14
2.2 Οι Βασικές Αρχές των Ανοιχτών Δημοσίων Δεδομένων.....	17
2.3 Γιατί θα πρέπει τα Δημόσια Δεδομένα να είναι Ανοιχτά ;.....	18
2.4 Οφέλη.....	19
2.4.1 Επαναχρησιμοποίηση.....	20
2.4.2 Διαλειτουργικότητα.....	22
2.4.3 Ελληνικό Πλαίσιο Διαλειτουργικότητας.....	23
2.4.4 Επίπεδα Διαλειτουργικότητας.....	25
2.5 Μεταδεδομένα.....	27
2.5.1 Πρότυπα Μεταδεδομένων.....	28
2.6 Πηγές Ανοιχτών Δεδομένων.....	30
ΚΕΦΑΛΑΙΟ 3 - Σημασιολογικός Ιστός , Συνδεδεμένα Δεδομένα και Συνδεδεμένα Κυβερνητικά Δεδομένα.....	37
3.1 Από τον Ιστό των Εγγράφων στον Ιστό των Δεδομένων.....	37
3.2 Κανόνες Συνδεδεμένων Δεδομένων.....	41
3.3 Θέματα Αρχιτεκτονικής Συνδεδεμένων Δεδομένων.....	42
3.3.1 Πληροφοριακοί και μη-Πληροφοριακοί Πόροι.....	45
3.3.2 Dereferencing HTTP URI's.....	47
3.3.3 Διαπραγμάτευση Περιεχομένου (content negotiation).	48
3.3.4 Ταυτόσημα URI's (URI aliases).....	51
3.3.5 Σχετιζόμενες Περιγραφές.....	51
3.4 Linking Open Data Project.....	52
3.5 Συνδεδεμένα Κυβερνητικά Ανοιχτά δεδομένα.....	56
3.5.1 Δημοσίευση Κυβερνητικών Δεδομένων στον Ιστό ως Συνδεδεμένα Δεδομένα	56

ΚΕΦΑΛΑΙΟ 4 - Οπτικοποίηση Δεδομένων	62
4.1 Τι είναι η Οπτικοποίηση Δεδομένων ;	62
4.2 Ιστορική Αναδρομή	63
4.3 Γραφήματα Πληροφοριών (infographics).....	65
4.4 Εξόρυξη Πληροφορίας	67
4.5 Οπτικοποίηση Ανοικτών Δεδομένων	69
4.5.1 Περιπτώσεις Χρήσης.....	71
4.5.2 Εργαλεία Οπτικοποίησης Δεδομένων.....	79
4.6 Κριτική Ανάλυση.....	81
ΚΕΦΑΛΑΙΟ 5 - Η Εφαρμογή.....	83
5.1. Στόχος	83
5.2. Προδιαγραφές.....	84
5.3. Διαύγεια	84
5.4. Υλοποίηση	87
5.4.1 Τεχνολογίες.....	87
5.4.2 Πραγμάτωση.....	88
5.4.3 Περιπτώσεις Χρήσης.....	90
ΚΕΦΑΛΑΙΟ 6 – Συμπεράσματα	101
6.1 Σύνοψη - Συμπεράσματα	101
6.2 Πηγές - Βιβλιογραφία	102

ΚΕΦΑΛΑΙΟ 1 - Εισαγωγή

Το διαδίκτυο αποτελεί ένα αναπόσπαστο κομμάτι της καθημερινότητας σχεδόν κάθε χρήστη ηλεκτρονικού υπολογιστή. Ο κάθε χρήστης χρησιμοποιεί το διαδίκτυο για να καλύψει κάποιες προσωπικές του ανάγκες. Αυτές μπορεί να είναι η ανάγκη για ενημέρωση, για ψυχαγωγία, για επικοινωνία κ.ά. Πολύ συνηθισμένες εφαρμογές του διαδικτύου είναι ο διαμοιρασμός αρχείων, η αλληλεπίδραση και η συνεργασία με άλλους χρήστες, η εκμετάλλευση του διαδικτύου για θέματα δουλειάς, η ανάπτυξη σχέσεων με άλλους χρήστες (κοινωνικά δίκτυα), η επιχειρηματικότητα κ.ά.

Σε κάθε εφαρμογή που σχετίζεται με το διαδίκτυο παράγονται και κάποια δεδομένα, των οποίων μάλιστα το μέγεθος ποικίλει. Μπορεί να είναι από ελάχιστα μικρό, όπως μια ανανέωση σε ένα κοινωνικό δίκτυο, έως και αρκετά μεγάλο, όπως η δημοσίευση μια γενετικής ακολουθίας. Την ίδια ποικιλομορφία παρουσιάζει και η χρησιμότητα που έχουν αυτά τα δεδομένα στους διάφορους χρήστες.



Όσο ο ηλεκτρονικός υπολογιστής διαδίδεται τόσο αυξάνονται και αυτοί που τον χρησιμοποιούν. Φυσικό επακόλουθο είναι να αυξάνονται και οι χρήστες του διαδικτύου. Αλυσιδωτά αυξάνονται και τα δεδομένα που παράγονται κάθε χρονική στιγμή. Και δεν είναι μόνο τα καινούργια δεδομένα που παράγονται κατά την διάρκεια χρησιμοποίησης του διαδικτύου, αλλά γίνονται διαθέσιμα και αυτά που ήδη έχουν παραχθεί και δεν υπήρχε πριν τρόπος διαμοιρασμός τους. Όλα αυτά οδηγούν σε μια τεράστια αύξηση στα δεδομένα που είναι διαθέσιμα φτάνοντας σε ένα σημείο κορεσμού όπου ο όγκος των δεδομένων είναι πολύ μεγαλύτερος από αυτόν που μπορούμε να αναλύσουμε.

Όλα αυτά τα δεδομένα μπορεί να είναι ελεύθερα προσβάσιμα από τον καθένα ή να ανήκουν σε κάποιες εταιρίες και να τα εκμεταλλεύονται για κερδοσκοπικούς λόγους (π.χ. χαρακτηριστικά χρηστών που χρησιμοποιούνται για θεωρητικά στοχευόμενη διαφήμιση) . Βέβαια οι σύγχρονες διαδικτυακές τάσεις προστάζουν όλα τα δεδομένα να είναι ελεύθερα προς επεξεργασία από τον καθένα, χωρίς κάποιο αντίτιμο.

Η προέλευση των δεδομένων μπορεί να ποικίλει και να είναι από απλούς χρήστες μέχρι και μεγάλες εταιρίες. Τα αποτελέσματα από την επεξεργασία είναι ανάλογα της φύσης των δεδομένων. Από τα πιο ενδιαφέροντα σύνολα δεδομένων είναι αυτά που προέρχονται από τις κυβερνήσεις. Η επεξεργασία τους οδηγεί στην ισχυροποίηση της συμμετοχής των πολιτών στα κοινά, κάτι που αποφέρει αρκετά θετικά στοιχεία στο κοινωνικό σύνολο. Για αυτό το λόγο οι κυβερνήσεις έχουν προβεί στην δημοσίευση των δεδομένων τους και στην χορήγηση άδειας για ελεύθερη εκμετάλλευση. Τα ανοιχτά κυβερνητικά δεδομένα βοηθούν και στην ορθή λειτουργία και επιτυχία της προσπάθειας για ηλεκτρονική διακυβέρνηση, μια προσπάθεια που έχουν ξεκινήσει οι κυβερνήσεις ανά τον κόσμο.

Η μεγάλη αύξηση στα δεδομένα δημιουργεί αρκετά προβλήματα για τα οποία είναι αναγκαίο να βρεθεί άμεσα λύση. Το βασικότερο πρόβλημα είναι η δυσκολία κατανόησης της πληροφορίας που κρύβουν τα μεγάλα σε όγκο δεδομένα. Σε αυτό το πρόβλημα την λύση δίνει η οπτικοποίηση δεδομένων, που είναι και το βασικό αντικείμενο μελέτης αυτής της εργασίας

Στην συνέχεια της εργασίας αυτής θα εξετάσουμε τα ανοιχτά διαδικτυακά δεδομένα (Κεφάλαιο 2), έπειτα θα επικεντρωθούμε στα Συνδεδεμένα Δεδομένα και Συνδεδεμένα Κυβερνητικά Δεδομένα (Κεφάλαιο 3). Ακολούθως πως μπορούμε να εκμεταλλευτούμε τα δεδομένα μέσω της οπτικοποίησης, προβάλλοντας αρκετές προσπάθειες οπτικοποίησης κυβερνητικών δεδομένων αλλά και διάφορα εργαλεία που μπορούν να χρησιμοποιηθούν για οπτικοποίηση (Κεφάλαιο 4).

Τέλος παρουσιάζουμε την εφαρμογή που υλοποιήσαμε (<http://greekopendata.host22.com/>), βασισμένου σε ελληνικά ανοιχτά δεδομένα, που αποτελεί το αποτέλεσμα όσων μελετήθηκαν (Κεφάλαια 5) και κλείνοντας συνοψίζουμε και αναφερόμαστε στα μελλοντικά σχέδια (Κεφάλαιο 7).

ΚΕΦΑΛΑΙΟ 2 - Ανοικτά Δημόσια Δεδομένα

Η δημόσια διοίκηση παράγει συλλέγει καθημερινά σύνολα δεδομένων, τα οποία αφορούν την οικονομική και κοινωνική ζωή των πολιτών. Η ελεύθερη διάθεση των δεδομένων αυτών υποστηρίζεται όλο και πιο συχνά τόσο από διεθνή παραδείγματα όσο και από τον δημόσιο διάλογο που διεξάγεται. Βρισκόμαστε ακόμα στις απαρχές, σε όρους κλιμάκωσης και δημοσιότητας, του κινήματος για τα Ανοικτά Δημόσια Δεδομένα. Ωστόσο, εδώ και μερικά χρόνια, το αίτημα λαμβάνει ολοένα και μεγαλύτερη προσοχή σε σχέση με το παρελθόν ενώ αξιόλογες πρωτοβουλίες έχουν ήδη δρομολογηθεί τόσο σε τοπικό, όσο και σε εθνικό και υπερεθνικό επίπεδο.

Σύμφωνα με εκτιμήσεις, μόνο ένα πολύ μικρό ποσοστό των δεδομένων που παράγουν οι κυβερνητικές υπηρεσίες και οργανισμοί, τελικά επαναχρησιμοποιούνται. Η αξιοποίηση των επικοινωνιακών και τεχνολογικών δυνατοτήτων που προσφέρει το διαδίκτυο, δημιουργεί νέες προοπτικές για την αξιοποίηση των δημοσίων δεδομένων που αν συνδυαστούν με πιο ανοιχτές πολιτικές πρόσβασης μπορούν να δώσουν μια δυναμική στην κοινωνία της πληροφορίας.

Το άνοιγμα των δεδομένων συμβάλλει στην επίτευξη τουλάχιστον τριών διαφορετικών στόχων: (α) στην διαφάνεια της διοίκησης και την ενίσχυση της δημοκρατικής συμμετοχής, (β) στην ενίσχυση της καινοτομίας και της γνώσης, και (γ) στην καλύτερη λειτουργία της ίδιας της διοίκησης. Η συμβολή του ανοίγματος των δημοσίων δεδομένων στην επίτευξη των στόχων αυτών, εξετάζεται αναλυτικά σε επόμενα κεφάλαια. Προτού εξετάσουμε όμως τις δυνατότητες και τις προοπτικές που ανοίγονται, παραθέτουμε κάποιες βασικές διευκρινιστικές σκέψεις και ορισμούς που θα φανούν χρήσιμοι στη συνέχεια.

2.1 Ορισμοί ¹

Στη δημόσια διαβούλευση και συζήτηση σχετικά με τα δεδομένα, εμφανίζονται και χρησιμοποιούνται οι ακόλουθοι επιθετικοί προσδιορισμοί : **Ανοιχτά** , **Δημόσια** , **Κυβερνητικά**. Η αποσαφήνιση των παραπάνω όρων καθώς και του πλαισίου εντός του οποίου χρησιμοποιούνται είναι σημαντική για την κατανόηση του θέματος που πραγματευόμαστε στη παρούσα μελέτη.

Στο σημείο αυτό κρίνεται σημαντικό να δώσουμε τον βασικό ορισμό της έννοιας «**δεδομένα**» και να τον διαχωρίσουμε από τον επίσης συχνά χρησιμοποιούμενο όρο «**πληροφορία**».

2.1.1 Δεδομένα και Πληροφορίες

Δεδομένα :

«Ως δεδομένα ορίζουμε 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 επιστήμη της πληροφορικής συναντούμε τα δεδομένα στον πληθυντικό αριθμό, σπανιότερα δε στον ενικό (δεδομένο)»

Πληροφορίες :

«Πληροφορία είναι το αποτέλεσμα επεξεργασίας των δεδομένων με στόχο την αύξηση της αξίας τους. Ως αξία εδώ νοείται η γνώση που μας προσφέρεται και η οποία βοηθάει στην κατανόηση της πραγματικότητας. Επομένως η συλλογή , ο συσχετισμός και η επεξεργασία των δεδομένων μας δίνουν τις πληροφορίες οι οποίες μας βοηθούν στην κατανόηση της πραγματικότητας.»

Τα δεδομένα δεν αναφέρουν τίποτα για τη σημαντικότητα τους ή τη σχέση τους προς οτιδήποτε. Σε αντίθεση με τα δεδομένα η πληροφορία έχει ορισμένο νόημα και είναι οργανωμένη για συγκεκριμένο σκοπό². Με την προσθήκη αξιών στα δεδομένα (με διαμόρφωση, οργάνωση, μαθηματική ή στατιστική ανάλυση, διόρθωση λαθών ή συμπύεση) παράγεται η πληροφορία. Τη διαδικασία αυτή ονομάζουμε **επεξεργασία δεδομένων**.

¹ Παπαδόπουλος Θεοδωρής. (2011) . Ανοιχτά Δημόσια Δεδομένα , Εθνική Σχολή Δημόσιας Διοίκησης , <http://thodoris.net/files/OpenPublicData.pdf>

Παρά την ύπαρξη των διακριτών αυτών ορισμών για τα δεδομένα και τις πληροφορίες, οι δύο αυτοί όροι χρησιμοποιούνται συχνά ο ένας στη θέση του άλλου. Ανάλογα με το «γνωσιολογικό ύφος» (cognitive style) και το υπόβαθρο αναφοράς, συναντάμε συχνά τον όρο δεδομένα να αναφέρεται σε οργανωμένα σύνολα που έχουν υποστεί επεξεργασία και εμπεριέχουν ποσότητα γνώσης, ενώ άλλες φορές ο όρος πληροφορία χρησιμοποιείται για να γίνει αναφορά σε μη επεξεργασμένα σύνολα πρωτογενών δεδομένων. Σε νομικά κείμενα προτιμάται ο όρος πληροφορία, ακόμα και για τις περιπτώσεις που δεν γνωρίζουμε το πληροφοριακό δυναμικό που εμπεριέχεται στον υπό αναφορά όρο.

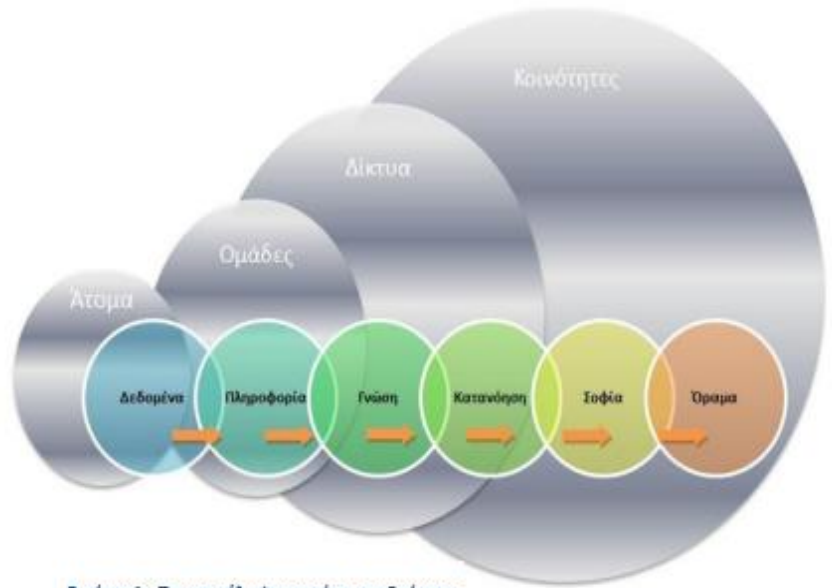
Στην μελέτη αυτή χρησιμοποιούμε συχνότερα τον όρο δεδομένα. Η επιλογή αυτή κρίθηκε ορθότερη τόσο γιατί οι όροι δεδομένα και ανοιχτά δεδομένα συναντούνται συχνότερα στην σχετική βιβλιογραφία όσο και επειδή η χρήση του όρου πληροφορία εισάγει έναν περιορισμό στα σύνολα δεδομένων που μπορούμε να αναφέρουμε. Σημειώνεται πάντως πως σε αρκετές περιπτώσεις οι δύο όροι χρησιμοποιούνται εναλλακτικά, ανάλογα με το γενικότερο πλαίσιο περιγραφής. Έτσι για παράδειγμα στις ενότητες που αναλύεται η νομική πτυχή των ανοιχτών δημοσίων δεδομένων χρησιμοποιούμε συχνά τον όρο «πρόσβαση στη δημόσια πληροφορία» καθώς αυτός αποτελεί την συνήθη ορολογία σε νομικά κείμενα.

2.1.2 Το Θεμελιώδες τετράπτυχο «Δεδομένα -Πληροφορία-Γνώση – Σοφία»

Αν τα δεδομένα δεν αναφέρουν τίποτα για τη σημαντικότητα τους ή τη σχέση τους προς οτιδήποτε, τότε γιατί τα μελετάμε και γιατί γίνεται τόσος θόρυβος γύρω από την προσπάθεια ανοίγματος των δημοσίων δεδομένων;

Ο βασικός λόγος είναι ότι τα δεδομένα αποτελούν το πρωταρχικό επίπεδο σε μια ιεραρχία που οδηγεί στην γνώση για τον κόσμο. Το θεμελιώδες τετράπτυχο «δεδομένα – πληροφορία – γνώση – σοφία» παρουσιάστηκε πρώτη φορά με συστηματικό τρόπο ως «Πυραμίδα ή Ιεραρχία της Γνώσης» (Zeleny 1987) και σήμερα είναι γνωστό με το αρκτικόλεξο DIKW (Data Information Knowledge and Wisdom Hierarchy).

Το μοντέλο της Ιεραρχίας της Γνώσης έχει επεκταθεί από πολλούς και έχει συνδυαστεί με διάφορα άλλα μοντέλα. Για παράδειγμα η εκδοχή που παρουσιάζεται στην εικόνα 1 εξυπηρετεί στην κατανόηση της διαδικασίας δημιουργίας της γνώσης σε σχέση με την αυξανόμενη ενσωμάτωση που αυτή επιφέρει. Τα στάδια από τα δεδομένα ως την σοφία, είναι τα ακόλουθα :



Εικόνα 1 : Το μοντέλο Ιεραρχίας της Γνώσης

1. Η ανακάλυψη μοτίβων σε σύνολα δεδομένων οδηγεί στην **πληροφορία**
2. Η βιωματική ανακάλυψη μοτίβων σε δομημένη πληροφορία ενδυναμώνει τη **γνώση**.
3. Στη πορεία του χρόνου η ερμηνεία της πραγματικότητας με βάση την αποκτηθείσα γνώση ενδυναμώνει την **κατανόηση**.
4. Η ηθική αξιολόγηση της γνώσης σε συνδυασμό με την κατανόηση αποκρυσταλλώνεται σε **σοφία** προσωπική ή συλλογική.
5. Η συλλογική σοφία μπορεί να οδηγήσει στην δημιουργία ενός **οράματος** για το μέλλον , που σηματοδοτεί μια κρίσιμη αλλαγή συλλογικής κατεύθυνσης.

Επομένως , τα δεδομένα και οι πληροφορίες αποτελούν τον φορέα διακίνησης της γνώσης , βάσει της οποίας αναπτύσσονται η κατανόηση και η σοφία και τελικά γεννιούνται τα οράματα που αλλάζουν ή όχι την πραγματικότητα

2.1.3 Ανοικτά Δημόσια Δεδομένα

«Ο όρος **Ανοικτά Δημόσια Δεδομένα (Open Public Data)** χρησιμοποιείται για να αναφερθούμε σε δεδομένα ή σύνολα δεδομένων που αφορούν το συλλογικό γίνεσθαι και για τα οποία υφίσταται μια συνειδητή και συνεπής πολιτική η οποία επιτρέπει την ελεύθερη διάθεση και επαναχρησιμοποίηση τους.»

Ο παραπάνω ορισμός δεν είναι ούτε πλήρης ούτε καθολικός καθότι τόσο η έννοια Ανοικτά όσο και ο προσδιορισμός Δημόσια , διαφοροποιούνται ανάλογα με το γενικότερο πλαίσιο περιγραφής και το νομικό σύστημα εντός του οποίου εξετάζονται. Στη συνέχεια θα προσπαθήσουμε να προσεγγίσουμε καλύτερα τους δύο αυτούς όρους , ξεκινώντας από τον προσδιορισμό της κλάσης των δημοσίων δεδομένων και συνεχίζοντας με τις προϋποθέσεις που εξασφαλίζουν την ανοιχτότητα τους.

- **Δημόσια Δεδομένα**

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.

Ποιτικά, 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 λόγος 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.

- **«Ανοιχτά» Δεδομένα**

Σε γενικές γραμμές για να μπορούν τα δημόσια δεδομένα να θεωρηθούν ανοιχτά θα πρέπει να είναι δυνατή η περαιτέρω χρήση και επαναδιάθεση τους. Ως περαιτέρω χρήση νοείται η επαναχρησιμοποίηση δεδομένων και πληροφοριών για σκοπό άλλο από αυτόν για τον οποίο παρήχθησαν, είτε αυτός είναι εμπορικός είτε όχι. Ο βαθμός της ανοιχτότητας των δεδομένων προσδιορίζεται από την ύπαρξη ή όχι περιορισμών στις παραπάνω δυνατότητες. Σύμφωνα με τον ορισμό της Ανοιχτότητας που δίνει το Open Knowledge Foundation (OKFN)²:

«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 στο δημιουργό και την παρόμοια χρήση».

Στον παραπάνω ορισμό:

(α) Ορίζονται οι **ελάχιστες δυνατότητες** που έχει οποιοσδήποτε χρήστης επί των δεδομένων, δηλαδή: η δυνατότητα χρησιμοποίησης, επαναχρησιμοποίησης και επαναδιάθεσης και

(β) Θέτονται τα **όρια των περιορισμών** που επιτρέπεται να ασκηθούν επί αυτών των δυνατοτήτων, δηλαδή: θ υποχρέωση αναφοράς του δημιουργού και/ι παρόμοια χρήση.

² <http://okfn.org/>

Οι δυνατότητες που αναφέρονται στο πρώτο κομμάτι του ορισμού θα πρέπει να διασφαλίζονται με δύο τρόπους. Κατά πρώτον νομικά : η άδεια που συνοδεύει τα δεδομένα θα πρέπει να επιτρέπει όλες τις παραπάνω χρήσεις. Κατά δεύτερον τεχνολογικά, θα πρέπει δηλαδή οι τεχνολογίες και τα τεχνολογικά μέσα που χρησιμοποιούνται , να επιτρέπουν και να διευκολύνουν τόσο την επαναχρησιμοποίηση όσο και την επαναδιάθεση.



Εικόνα 2 : Το οικοσύστημα των Ανοιχτών Δεδομένων

2.2 Οι Βασικές Αρχές των Ανοιχτών Δημοσίων Δεδομένων

Τα κυβερνητικά δεδομένα θα θεωρούνται ανοιχτά, αν διατίθενται με τρόπο που συνάδει με τις παρακάτω αρχές³ :

1. **Πλήρη.** Όλα τα δημόσια δεδομένα διατίθενται χωρίς περιορισμό. Δημόσια είναι όλα τα δεδομένα που δε υπόκεινται σε αναγκαίο περιορισμό λόγω προσωπικής φύσης ή διαβάθμισης
2. **Πρωτογενή δεδομένα** είναι αυτά που συλλέγονται στην πηγή της δημιουργίας του και διαθέτουν τον υψηλότερο βαθμό κατάτμησης (κβαντοποίησης) χωρίς αλλαγές ή συγχωνεύσεις.
3. Τα δεδομένα **δημοσιεύονται όσο το δυνατό γρηγορότερα** ώστε να διατηρούν την χρησιμότητά τους.
4. Τα δεδομένα είναι **ευρέως διαθέσιμα** στο μεγαλύτερο δυνατό εύρος χρηστών και χρησιμοτήτων.
5. Τα δεδομένα θα πρέπει να βρίσκονται σε **μορφή φιλική προς υπολογιστική επεξεργασία από την υπάρχουσα τεχνολογία.**
6. **Δεδομένα χωρίς διακρίσεις.** Τα δεδομένα είναι διαθέσιμα σε όλους/ες του ενδιαφερόμενους/ες χωρίς την ανάγκη εγγραφής προσωπικών δεδομένων.
7. **Ανοιχτά δεδομένα.** Τα δεδομένα θα πρέπει να διατίθενται σε μορφή της οποίας καμία οντότητα δεν θα έχει αποκλειστικό έλεγχο ή πνευματικά δικαιώματα.
8. **Αποποίηση πνευματικών δικαιωμάτων.** Τα δεδομένα δεν θα δεσμεύουν πνευματικά δικαιώματα, πατέντες, λογότυπα κτλ. Λογικοί περιορισμοί θα μπορούν να επιβληθούν σε περιπτώσεις ανάγκης προσωπικής προστασίας δεδομένων, ασφάλειας ή διαβάθμισης.
9. Τα δεδομένα που εισέρχονται σε αυτές τις κατηγορίες θα πρέπει να είναι **επιθεωρίσιμα** ⁴.

³ Το κείμενο αποτελεί μετάφραση από το αυθεντικό κείμενο τοποθέτησης όπως είχε παρουσιαστεί το 2007 από επιτροπή υπό τους Tim O'Reilly και Carl Malamud (Open Government Working Group - https://public.resource.org/open_government_meeting.html)

⁴ “επιθεωρίσιμα” σημαίνει ότι θα πρέπει να οριστεί ένα άτομο επικοινωνίας όταν το κοινό προσπαθεί να χρησιμοποιήσει τα δεδομένα. Θα πρέπει να οριστεί ένα άτομο επικοινωνίας όταν γίνονται παράπονα σχετικά με παραβιάσεις των αρχών χρήσης. Μία δικαστική αρχή θα πρέπει να έχει τη δικαιοδοσία να επιθεωρήσει αν η υπηρεσία διαχείρισης των δεδομένων έχει εφαρμόσει τις αρχές ορθά.

Το σημείο κλειδί έγκειται στο ότι, όταν «ανοίγει» κάποιος τα δεδομένα, στο επίκεντρο πρέπει να βρίσκονται μόνο μη-προσωπικά δεδομένα, δηλαδή δεδομένα που δεν περιέχουν πληροφορίες για συγκεκριμένα άτομα. Όμοια, για ορισμένα είδη κυβερνητικών δεδομένων, μπορεί να ισχύουν περιορισμοί εθνικής ασφάλειας.

2.3 Γιατί θα πρέπει τα Δημόσια Δεδομένα να είναι Ανοιχτά ⁵;

Τα Ανοιχτά Δεδομένα πρέπει να είναι ανοιχτά τόσο από νομικής άποψης, όσο και από τεχνικής. Ειδικότερα, τα δεδομένα πρέπει να είναι διαθέσιμα αυτούσια στο σύνολο τους, αλλά και σε μορφή αναγνωρίσιμη από τις μηχανές.

Υπάρχουν πολλοί λόγοι για τους οποίους τα δημόσια δεδομένα θα πρέπει να είναι ανοιχτά προς την κοινωνία. Καταρχήν επειδή τα δεδομένα αυτά παράγονται με χρήματα των φορολογούμενων πολιτών που χρηματοδοτούν έρευνες και υπηρεσίες της Κυβέρνησης και συνεπώς αποτελεί αναφαίρετο δικαίωμα των πολιτών να έχουν δυνατότητα πρόσβασης στα αποτελέσματα του έργου αυτού.

Δεύτερον γιατί τα δεδομένα αυτά αν συνδυαστούν και μετασηματιστούν κατάλληλα εμπεριέχουν πολύτιμη γνώση, για την κοινωνική και οικονομική πραγματικότητα, που κα πρέπει να καθίσταται διαθέσιμη και να υποστηρίζει τη διαδικασία των συλλογικών αποφάσεων.

Τρίτον γιατί μέρος αυτών των δεδομένων αποτυπώνει και καταγράφει την αποτελεσματικότητα της δημόσιας ασκούμενης πολιτικής και άρα αποτελεί παράγοντα αξιολόγησης του έργου της Κυβέρνησης. Σε μία απώτερη αλλά ορατή αναγωγή, τα ανοιχτά κυβερνητικά δεδομένα μειώνουν την απόσταση μεταξύ πολίτη και εξουσίας, αφού αν αξιοποιηθούν κατάλληλα οδηγούν σε πληροφορίες και γνώση, βάσει της οποίας οι πολίτες τροφοδοτούν με σήματα το πολιτικό σύστημα μέσω της ψήφου τους.

Επιπροσθέτως, τα ανοιχτά και ελεύθερα διαθέσιμα δημόσια δεδομένα ενισχύουν την αποτελεσματικότητα του δημόσιου τομέα μεταφέροντας κάποιες από τις αναλυτικές απαιτήσεις της διοίκησης σε τρίτα μέρη όπως κοινότητες πολιτών, μη κυβερνητικές οργανώσεις, ερευνητικά

⁵ Παπαδόπουλος Θεοδωρής. (2011). Ανοιχτά Δημόσια Δεδομένα, Εθνική Σχολή Δημόσιας Διοίκησης, <http://thodoris.net/files/OpenPublicData.pdf>

ιδρύματα κ μέσα μαζικής ενημέρωσης, τα οποία συνδυάζουν δεδομένα από διαφορετικές πηγές με πρωτότυπο και εφευρετικό τρόπο.

2.4 Οφέλη

Τα κυριότερα οφέλη των ανοιχτών κυβερνητικών δεδομένων αποτυπώνονται, καταρχήν, από τα χαρακτηριστικά τους:

- είναι ανοιχτά, άρα διαθέσιμα ανά πάσα στιγμή και σε οποιαδήποτε τοποθεσία χρήσης του διαδικτύου και εξαλείφει την ανάγκη φυσικής παρουσίας του τελικού χρήστη σε διοικητικά γραφεία και την πολύωρη αναμονή σε ουρές
- είναι κυβερνητικά, άρα ο πολίτης μπορεί να ενημερωθεί για την εφαρμογή της κυβερνητικής πολιτικής,
- είναι αξιόπιστα, καθώς μπορεί να πιστοποιηθεί η προέλευση τους,
- είναι ψηφιοποιημένα, αυτό σημαίνει μικρότερο χρόνο διαχείρισής τους (αναζήτησης, προσπέλασης, χρήσης).

Σημαντικό όφελος αποτελεί ο περιορισμός και η πλήρης εξάλειψη της **γραφειοκρατίας** με στόχο τη βελτίωση της ποιότητας ζωής των πολιτών και την αρωγή στην ομαλή λειτουργία μεταξύ Δημόσιας Διοίκησης και Πολιτών. Επίσης, τονώνεται η επιχειρηματικότητα, αφού εύλογα η γραφειοκρατία αποτελεί πολλές φορές ανασταλτικό παράγοντα για την ανάπτυξη μιας επιχείρησης.

Το **οικονομικό όφελος** είναι το πρώτο ζήτημα της σημερινής εποχής. Σε πρόσφατη πτυχιακή εργασία^[22], αποδείχτηκε ότι το προσδοκώμενο κέρδος από το μετασχηματισμό και την πλήρη ηλεκτρονική ολοκλήρωση τριών υπηρεσιών του ΔΔ κυμαίνεται μεταξύ 88 και 92 %. Σύμφωνα με τις εκτιμήσεις, η οικονομική ωφέλεια της εφαρμογής εκτιμάται στα 4 δις ευρώ. Βέβαια, η οικονομική ωφέλεια δεν αποδίδεται μόνο στην δημοσιοποίηση των δεδομένων των παραπάνω υπηρεσιών αλλά στη συνολική παροχή ηλεκτρονικής υπηρεσίας, δεν παύουν όμως τα δεδομένα, ως βήμα εκκίνησης, να προσδίδουν μεγάλη αξία στην υπηρεσία αυτή. Επίσης, είναι βολική και κερδοφόρα για τις επιχειρήσεις, όπως και επωφελής για το κοινό, η εύκολη πρόσβαση σε πρόσφατες πληροφορίες χωρίς να πρέπει να ξοδεύονται χρόνος, ενέργεια και χρήματα.

Η **διαχείριση** της πληροφορίας είναι ένα σημαντικό θέμα. Καταρχήν η πρόσβαση σε κυβερνητικές πληροφορίες είναι πιο εύκολη για τους πολίτες. Επίσης, οι νέες τεχνολογίες προσφέρουν τα

κατάλληλα εργαλεία διαχείρισης των δεδομένων όπως ταξινόμηση, κατηγοριοποίηση, αναζήτηση κα. ιδίως σε περιπτώσεις μεγάλου όγκου δεδομένων. Τα δεδομένα μπορούν να ενσωματωθούν σε συσκευές και υπηρεσίες που ο πολίτης χρησιμοποιεί καθημερινά όπως τα συστήματα γεωπληροφόρησης (GPS), μοντέλα πρόβλεψης καιρού, εφαρμογές λογισμικού σε κινητά κλπ.

Η προσφορά της εφαρμογής των ανοιχτών δημόσιων δεδομένων συνάδει με τους στόχους της ηλεκτρονικής διακυβέρνησης, όπως είναι η συμμετοχή των πολιτών στη διαμόρφωση της πολιτικής μιας χώρας. Μέσω του διαδικτύου άνθρωποι από όλη την χώρα μπορούν να αλληλοεπιδράσουν με πολιτικούς και να ακουστούν. Ηλεκτρονική διαβούλευση και ενημέρωση των πολιτών για τις τελευταίες πολιτικές και οικονομικές εξελίξεις (και μάλιστα κατευθείαν από την πηγή των δεδομένων) θα επιτρέψουν πολίτες να εκφράσουν σε δεύτερο βήμα τις απόψεις τους για τον τρόπο εφαρμογής της κυβερνητικής πολιτικής. Θέμα.

Οι τεχνολογίες αυτές μπορούν να δημιουργήσουν μια πιο διαφανή διακυβέρνηση. Τα ανοιχτά δεδομένα δίνουν τη δύναμη στον πολίτη και τις επιχειρήσεις να ενημερωθούν για τις εφαρμογές της πολιτείας και τα οικονομικά μεγέθη. Κατά συνέπεια, βοηθά τους ψηφοφόρους να σχηματίσουν μια ολοκληρωμένη γνώμη για το έργο της κυβέρνησης και να αποφασίζουν καλύτερα πώς να ψηφίσουν στο μέλλον. Από την άλλη πλευρά, οι κυβερνήσεις ενημερώνονται

άμεσα για τις πολιτικές των άλλων χωρών σε όλα τα επίπεδα, άρα προσφέρεται αβίαστα η γνώση και οι τεχνικές ώστε να μπορούν να εφαρμόσουν τη δική τους πολιτική αποδοτικότερα.

Η διαχείριση των δεδομένων είναι επίσης απαραίτητη για την αντιμετώπιση των περιβαλλοντικών προκλήσεων. Παραδείγματα αποτελούν η επεξεργασία των τρόπων κατανάλωσης ενέργειας για τη βελτίωση της ενεργειακής απόδοσης ή των δεδομένων που αφορούν τη ρύπανση για τη διαχείριση της κυκλοφορίας. Τεκμηριωμένες αποφάσεις χάραξης πολιτικής στους τομείς των μεταφορών, των χρήσεων γης και της κλιματικής αλλαγής εξαρτώνται όλο και περισσότερο από την ανάλυση των διαθέσιμων δεδομένων.

2.4.1 Επαναχρησιμοποίηση

Οι πληροφορίες του δημόσιου τομέα είναι η μεγαλύτερη ενιαία πηγή πληροφοριών στην Ευρώπη. Η πληροφορία παράγεται και συλλέγεται από δημόσιους φορείς και περιλαμβάνει ψηφιακούς χάρτες, μετεωρολογικές, νομικές, κίνηση, χρηματοπιστωτική, οικονομική και άλλα δεδομένα. Τα

περισσότερα από αυτά τα ανεπεξέργαστα δεδομένα θα μπορούσαν να επαναχρησιμοποιηθούν ή να ενσωματωθούν σε νέα προϊόντα και υπηρεσίες, τα οποία χρησιμοποιεί ο πολίτης σε καθημερινή βάση, όπως τα συστήματα πλοήγησης αυτοκινήτων, προβλέψεις καιρού, χρηματοπιστωτικές και ασφαλιστικές υπηρεσίες.

Παρατηρούνται εταιρείες να επαναχρησιμοποιούν δημόσια δεδομένα όπως δεδομένα που έχουν σχέση με πολεοδομικά στοιχεία για να δημιουργήσουν ή να διορθώσουν χάρτες και συστήματα πλοήγησης, εμπορικές εταιρείες να χρησιμοποιούν στατιστικά οικονομικά στοιχεία ώστε να πορευτούν για την πολιτική της παραγωγής τους ακόμα και αγρότες μετά από ανάλυση οικονομικών δεδομένων των εξαγωγών της χώρας να αποφασίζουν για την επιλογή των καρπών που φυτέψουν ώστε να έχουν περισσότερες πιθανότητες πώλησης των αγροτικών τους προϊόντων.

Συνεπώς, το σημαντικότερο όφελος των ανοιχτών δεδομένων και δη των δημοσίων είναι η **επαναχρησιμοποίησή** τους, δηλαδή η περαιτέρω εκμετάλλευση των δεδομένων αυτών ώστε να παραχθεί νέα πληροφορία. Ως επαναχρησιμοποίηση δεδομένων του δημόσιου τομέα θεωρείται ο συνδυασμός τους ώστε να παραχθεί νέα πληροφορία με προστιθέμενη αξία στα αρχικά δεδομένα και η αξιοποίησή τους σε νέες εφαρμογές, τόσο για εμπορικούς και μη εμπορικούς σκοπούς. Η περαιτέρω ανάλυσή των αρχέγονων δεδομένων θα εξάγει τους πολυπρόθετους δείκτες και συμπεράσματα που θα στηρίζονται στο συγκερασμό δεδομένων από διάφορες πηγές, με σκοπό την ενίσχυση της διαφάνειας και της αξιοπιστίας, τη δημιουργία κοινωνικής και οικονομικής αξίας και τη βελτίωση των δημοσίων υπηρεσιών.

Με αυτόν τον τρόπο, οι πληροφορίες του δημόσιου τομέα προσδίδουν μεγάλες οικονομικές δυνατότητες. Σύμφωνα με μια έρευνα σχετικά με τον οικονομικό αντίκτυπο των υφιστάμενων πληροφοριών του δημόσιου τομέα στην Ευρωπαϊκή Ένωση (ΕΕ) το 2011, (μελέτη Vickery[Υποσημείωση]) τα συνολικά (άμεσα και έμμεσα) οικονομικά κέρδη εκτιμώνται σε 140δισ ευρώ σε ολόκληρη την ΕΕ. Παρόμοια έρευνα το 2006 (η μελέτη Mespir[Υποσημείωση]) εκτιμούσε ότι η συνολική αγορά του PSI στην ευρωπαϊκή ένωση ανερχόταν στα 27 δισ ευρώ. Συνακόλουθο της επαναχρησιμοποίησης είναι η δημιουργία νέων επιχειρήσεων (άρα και νέων θέσεων εργασίας) καθώς νέα προϊόντα και υπηρεσίες πληροφορίας προκύπτουν από την ανάλυση και την οπτικοποίηση των δεδομένων από διαφορετικές πηγές, γεγονός που συνεπάγει την ανάπτυξη της επιχειρηματικότητας.

Η επαναχρησιμοποίηση των δεδομένων που ανήκουν στο δημόσιο τομέα δε θα πρέπει να υπόκειται σε περιορισμούς ευρεσιτεχνίας. Ακόμα πιο σημαντικό, τα δεδομένα να είναι διαθέσιμα σε μορφή αναγνωρίσιμη από τις μηχανές, κάτι το οποίο επιτρέπει ακόμα μεγαλύτερη επαναχρησιμοποίηση. Συγκεκριμένα, έστω ότι υπάρχουν κάποια στατιστικά στοιχεία δημοσιοποιημένα, σε μορφή PDF που χρησιμοποιείται συχνά για εκτυπώσεις υψηλής ποιότητας. Ενώ αυτά τα στατιστικά στοιχεία μπορούν να αναγνωστούν από τους ανθρώπους, καθίσταται πολύ δύσκολη η προσπέλαση τους από μία μηχανή.

Σήμερα είναι πια γνωστό ότι αυτή η συγκεντρωμένη πληροφορία, αλλά και η γνώση που μπορεί να αποκτηθεί από την επεξεργασία της, έχουν πολύ μεγάλη αξία. Προϋπόθεση για την αξιοποίηση της είναι η δυνατότητα των συστημάτων να επικοινωνούν μεταξύ τους. Είναι λοιπόν απαραίτητα τα συστήματα που σχεδιάζονται και υλοποιούνται να μπορούν να επικοινωνούν και να ανταλλάσσουν δεδομένα μεταξύ τους (αρχή της **διαλειτουργικότητας**).

2.4.2 Διαλειτουργικότητα

Τα πληροφορικά συστήματα, γενικά, βοηθούν τις κυβερνήσεις να επιτύχουν ανάπτυξη των οικονομιών τους, αλλά και να αντιμετωπίσουν την πολυπλοκότητα της παγκοσμιοποίησης των οικονομιών, τις απαιτήσεις των πολιτών και τις όποιες χρηματοοικονομικές πιέσεις. Για να τα διαχειριστούν όλα αυτά, χρειάζεται μεγάλη ευελιξία από την πλευρά των κυβερνήσεων στον τρόπο με τον οποίο διαχειρίζονται την πληροφορία. Είναι αναγκαίο όλα τα διαθέσιμα συστήματα της Δημόσιας Διοίκησης να επικοινωνούν με ενιαίο, ομοιόμορφο τρόπο μεταξύ τους. Ταυτόχρονα, χρειάζεται να γίνεται εύκολα η αναδιοργάνωσή τους και απαιτείται ευελιξία πρόσβασης στην τεχνολογία από ποικιλία προμηθευτών, αλλά και αξιοποίηση της πρωτοποριακής ανερχόμενης τεχνολογίας. Όλα αυτά γίνονται δυνατά χάρη στην αρχή της διαλειτουργικότητας.

Στην πράξη, **διαλειτουργικότητα** είναι η ικανότητα των συστημάτων πληροφορικής στο να παρέχουν ανοιχτά, σαφή και προτυποποιημένα σημεία επικοινωνίας. Όταν αυτά τα σημεία επικοινωνίας είναι ανοικτά τότε γίνεται και πιο εύκολη η βελτίωση και αντικατάσταση τους, παράγοντας πραγματικά ολοκληρωμένες υπηρεσίες.

Η διαλειτουργικότητα παρέχει πολλαπλά πλεονεκτήματα σε επιχειρήσεις, την κυβέρνηση και την ευρύτερη οικονομία μέσα από το ηλεκτρονικό επιχειρείν (e-business).

2.4.3 Ελληνικό Πλαίσιο Διαλειτουργικότητας

Στηριζόμενο στην ευρωπαϊκή πρωτοβουλία, έχει αναπτυχθεί ανάλογο Ελληνικό Πλαίσιο Διαλειτουργικότητας (ή Πλαίσιο Ηλεκτρονικής Διακυβέρνησης – ΠΗΔ) [10], δηλαδή ένα κείμενο προδιαγραφών για τεχνικά θέματα και προδιαγραφές, σύμφωνα με το οποίο η διαλειτουργικότητα ορίζεται ως η ικανότητα μεταφοράς και χρησιμοποίησης της πληροφορίας με ένα ομοιογενή και αποτελεσματικό τρόπο μεταξύ διαφόρων οργανισμών σε επίπεδο συστημάτων πληροφορικής. Το πλαίσιο διαλειτουργικότητας βασίστηκε στην αντίστοιχη διεθνή εμπειρία, έλαβε υπόψη τα αποτελέσματα σύντομης καταγραφής λειτουργικών αναγκών διαλειτουργικότητας ενός αριθμού δημόσιων φορέων και εντάσσεται στο συνολικό σχεδιασμό της Ελληνικής Δημόσιας Διοίκησης για την παροχή υπηρεσιών Ηλεκτρονικής Διακυβέρνησης σε φορείς, επιχειρήσεις και πολίτες.

Αποτελεί τον ακρογωνιαίο λίθο της Ψηφιακής Στρατηγικής 2006-2013[Υποσημείωση] για τη μετάβαση και προσαρμογή των υπηρεσιών της στις απαιτήσεις της σύγχρονης εποχής και είναι άμεσα συνυφασμένο με τους στόχους και τις κατευθύνσεις της Ευρωπαϊκής Πολιτικής 2010 – Ευρωπαϊκή Κοινωνία της Πληροφορίας 2010.

Το Πλαίσιο Ηλεκτρονικής Διακυβέρνησης στοχεύει στην αποτελεσματική υποστήριξη της Ηλεκτρονικής Διακυβέρνησης (e-governance) σε Κεντρικό, Περιφερειακό και Τοπικό επίπεδο και στη συμβολή στην επίτευξη της διαλειτουργικότητας σε επίπεδο πληροφοριακών συστημάτων, διαδικασιών και δεδομένων. Κάθε νέο σύστημα που πρόκειται να εγκατασταθεί σε Ελληνικό οργανισμό δημόσιου θα πρέπει απαραίτητα να είναι συμβατό με το Ελληνικό πλαίσιο διαλειτουργικότητας.

Οι βασικές αποφάσεις τεχνικής πολιτικής πάνω στις οποίες διαμορφώθηκε το ελληνικό πλαίσιο διαλειτουργικότητας είναι:

- Ευθυγράμμιση με τις υπηρεσίες διαδικτύου για όλα τα συστήματα πληροφορικής του δημοσίου τομέα
- Η υιοθέτηση της XML ως το κύριο πρότυπο για ολοκλήρωση δεδομένων και εργαλείων παρουσίασης για όλα τα συστήματα του δημοσίου τομέα
- Η υιοθέτηση του web browser ως το κύριο μέσο πρόσβασης του πολίτη στις υπηρεσίες της Ηλεκτρονικής Διακυβέρνησης
- Η προσθήκη των μεταδεδομένων (metadata) στο δυναμικό των κυβερνητικών πληροφοριών
- Η απαίτηση για συμμόρφωση με το πλαίσιο σε ολόκληρο τον δημόσιο τομέα.

Τέλος, η επιλογή των τεχνικών προδιαγραφών καθοδηγήθηκε από τα εξής:

- **Υποστήριξη από την αγορά** – οι προδιαγραφές που εκλέχθηκαν υποστηρίζονται ευρέως από την αγορά και κατά πάσα πιθανότητα θα μειώσουν το κόστος και το ρίσκο των κυβερνητικών συστημάτων πληροφορικής
- **Κλιμάκωση (Scalability)** - οι προδιαγραφές που εκλέχθηκαν έχουν την δυνατότητα να ικανοποιούν απαιτήσεις, όπως αλλαγές μεγέθους όγκου δεδομένων, αριθμού συναλλαγών και αριθμού χρηστών
- **Ανοικτά πρότυπα (Openness)** – Οι προδιαγραφές βασίζονται σε ανοικτά και ευρέως αποδεκτά πρότυπα που είναι καταγεγραμμένα και διαθέσιμα στο ευρύ κοινό.

Το Πλαίσιο Διαλειτουργικότητας Ηλεκτρονικής Διακυβέρνησης (ΠΔΗΔ) θέτει τις τεχνικές προδιαγραφές και πολιτικές και για την επίτευξη της διαλειτουργικότητας και συνοχής των συστημάτων πληροφορικής του δημοσίου τομέα, ορίζοντας τις βασικές προαπαιτήσεις για μία ολοκληρωμένη και ηλεκτρονική κυβέρνηση.

Ο άμεσος στόχος του πλαισίου διαλειτουργικότητας είναι να διευκολύνει την υλοποίηση της Ηλεκτρονικής Διακυβέρνησης και την συνολική υιοθέτησής της από ολόκληρο το δημόσιο τομέα.

Ο τελικός στόχος είναι δημόσιες υπηρεσίες που να ικανοποιούν τις ανάγκες επικοινωνίας με τον πολίτη, τους οργανισμούς και τις επιχειρήσεις μέσα από την αποτελεσματική και απρόσκοπτη ροή και προς τις δυο κατευθύνσεις.

Η υλοποίηση αυτού του οράματος βασίζεται στην ικανότητα των συστημάτων πληροφορικής του δημοσίου να παρέχουν σαφή και προτυποποιημένα σημεία επικοινωνίας (εναρμονισμένα με διεθνή και ευρωπαϊκά πρότυπα). Το πλαίσιο προτυποποιεί τα εξής:

- **Καθορισμένη μορφή πληροφοριών για ανταλλαγή** (μορφή πληροφορίας και δεδομένων). Δηλαδή καθορίζει τα γενικά πρότυπα δόμησης της πληροφορίας και της μετα-πληροφορίας που θα πρέπει να υιοθετηθούν από κάθε πληροφοριακό σύστημα του δημοσίου για να υποστηρίζουν δυνατότητα επικοινωνίας με άλλα συστήματα.
- **Καθορισμένος τρόπος ανταλλαγής πληροφοριών** (επικοινωνία / πρωτόκολλα). Καθορίζεται ένα σύνολο από τεχνολογίες επικοινωνίας που μπορούν να χρησιμοποιηθούν από συστήματα που θα είναι συμβατά με το ΠΔΗΔ, μέσω των οποίων θα υλοποιείται η

διαλειτουργικότητα (μεταφέρονται τα δεδομένα με την μορφή που αναφέρθηκε στο προηγούμενο σημείο).

- **Καθορισμένος τρόπος πρόσβασης πληροφοριών** (ασφάλεια / έλεγχος πρόσβασης). Το πλαίσιο προσδιορίζει και το σύνολο των τεχνολογιών που μπορούν να χρησιμοποιηθούν για να διασφαλίσουν την προστασία των υπηρεσιών διαλειτουργικότητας που θα προσφέρουν τα συμβατά συστήματα. Προφανώς, η ασφάλεια ενός πληροφοριακού συστήματος δεν εξαρτάται μόνο από την ασφάλιση του υποσυστήματος διαλειτουργικότητας.
- **Καθορισμένος τρόπος αναζήτησης πληροφοριών** (υπηρεσίες καταλόγου). Τέλος, το ΠΔΗΔ καθορίζει τις τεχνολογίες metadata και καταλόγου για την αναζήτηση των απαιτούμενων e-services και πληροφοριών στο πλήρως διαλειτουργικό μέλλον των πληροφοριακών συστημάτων που προτείνει.

2.4.4 Επίπεδα Διαλειτουργικότητας

Ανάλογα με το αν πρόκειται για επικοινωνία συστημάτων, οργανισμών, η ανθρώπων η διαλειτουργικότητα στην ηλεκτρονική διακυβέρνηση διακρίνεται σε θεσμική, τεχνική, σημασιολογική και οργανωτική διαλειτουργικότητα ^[23,24].

1. Θεσμική διαλειτουργικότητα

Αναφέρεται στην εναρμόνιση των νομοθετικών διατάξεων που διέπουν τη λειτουργία δύο ή περισσότερων φορέων που επιθυμούν να συνεργαστούν για τη μεταξύ τους ανταλλαγή πληροφοριών ή/ και την παροχή ολοκληρωμένων ηλεκτρονικών υπηρεσιών προς πολίτες, επιχειρήσεις και άλλους φορείς. Επιπλέον αποσκοπεί στο να διασφαλίσει ότι οι ηλεκτρονικά ανταλλασσόμενες πληροφορίες έχουν την ίδια νομική ισχύ για όλους τους εμπλεκόμενους φορείς. Η Θεσμική Διαλειτουργικότητα διασφαλίζεται μέσω νομοθετικών ρυθμίσεων και διατάξεων

2. Τεχνική διαλειτουργικότητα

Σε τεχνικό επίπεδο, διαλειτουργικότητα ορίζεται η ικανότητα μεταφοράς και χρησιμοποίησης της πληροφορίας με ομοιογενή και αποτελεσματικό τρόπο μεταξύ συστημάτων πληροφορικής

και οργανισμών.

Για την επίτευξη της διαλειτουργικότητας σε επίπεδο συστημάτων πληροφορικής, και οργανισμών, έχει δημιουργηθεί ευρωπαϊκό και ελληνικό πλαίσιο διαλειτουργικότητας για την ηλεκτρονική διακυβέρνηση. Αυτό το πλαίσιο διαλειτουργικότητας περιέχει τεχνικές προδιαγραφές και περιγράφει αναλυτικά τρόπους αποθήκευσης, μεταφοράς και παρουσίασης δεδομένων. Κάθε νέα εφαρμογή και πληροφοριακό σύστημα του δημοσίου τομέα θα πρέπει να είναι συμβατή με το Ελληνικό πλαίσιο διαλειτουργικότητας.

Διαλειτουργικά δίκτυα και βάσεις δεδομένων συντονίζονται σε περιφερειακό επίπεδο ώστε να είναι διαχειρίσιμα εύκολα και αποτελεσματικά. Οι τοπικοί φορείς παρακινούνται ώστε να δεχτούν κοινές λύσεις, με στόχο να αποφευχθεί η δαπανηρή αναπαραγωγή λογισμικού. Τέλος, παρέχεται ένα ομοιογενές περιβάλλον διασύνδεσης για ηλεκτρονικές υπηρεσίες στη βάση της περιφερειακής οργάνωσης.

3. Οργανωτική διαλειτουργικότητα

Η οργανωτική διαλειτουργικότητα αναφέρεται στον καθορισμό στόχων, τη διαμόρφωση διαδικασιών και την επίτευξη συνεργασίας των φορέων που επιδιώκουν ανταλλαγή πληροφοριών και ίσως έχουν διαφορετικές εσωτερικές δομές και διαδικασίες. Επιπλέον στοχεύει στην ικανοποίηση των απαιτήσεων της κοινότητας των χρηστών προσφέροντας υπηρεσίες αναγνωρίσιμες, προσβάσιμες και επικεντρωμένες στις ανάγκες του χρήστη.

4. Σημασιολογική διαλειτουργικότητα

Η Σημασιολογική διαλειτουργικότητα (Semantic Interoperability) αφορά τη διασφάλιση ότι η ακριβής έννοια-σημασία των ανταλλασσόμενων πληροφοριών είναι κατανοητή από οποιαδήποτε εφαρμογή. Έτσι επιτρέπει στα συστήματα να συνδυάζουν τις πληροφορίες με εκείνες από άλλες πηγές και να τις επεξεργάζονται αποτελεσματικά.

Η σημασιολογική διαλειτουργικότητα επιτυγχάνεται ορίζοντας και υιοθετώντας κοινό λεξιλόγιο και ορολογία σε όλα τα συστήματα και υπηρεσίες. Ο ορισμός και η συντήρηση ενός τέτοιου «λεξικού» γίνεται συνήθως από μια κεντρική υπηρεσία.

Η παρούσα εργασία προσανατολίζεται στη σημασιολογική διαλειτουργικότητα καθώς τα ανοιχτά δεδομένα πρέπει να ακολουθούν συγκεκριμένους κανόνες κατά τη δημοσιοποίησή τους ώστε να είναι επαναχρησιμοποιήσιμα. Στο σημείο αυτό εισάγεται η έννοια των μεταδεδομένων.

Οι τεχνολογικές εξελίξεις παρέχουν πρόσφορο έδαφος για την ανάπτυξη αναβαθμισμένων ποιοτικά συστημάτων καθώς παρουσιάζεται η ανάγκη διασύνδεσης όλων των δημόσιων υπηρεσιών για άμεση και αποτελεσματική ροή δεδομένων και ανταλλαγή πληροφοριών εξοικονομώντας χρόνο εργασίας και έρευνας. Η διασύνδεση υπηρεσιών της Δημόσιας Διοίκησης θέτει ζητήματα διαλειτουργικότητας, αρχή η οποία διατερείται με την ανάπτυξη πληροφοριακών συστημάτων που υλοποιούν διεθνή πρότυπα για την περιγραφή και κωδικοποίηση των διοικητικών ηλεκτρονικών εγγράφων. Συνεπώς απαιτείται η υιοθέτηση κοινού προτύπου περιγραφής (θεματικής και βιβλιογραφικής) για την κωδικοποίηση των δεδομένων με στόχο τη δυνατότητα επικοινωνίας και ανταλλαγής μηχαναγνώσιμων δεδομένων.

2.5 Μεταδεδομένα

Η δομημένη περιγραφή των δεδομένων βάση ενός προτύπου καλείται μεταδεδομένο (το πρόθεμα «μετα» διακιοιολογείται από τον ορισμό «δεδομένα για τα δεδομένα»). Για παράδειγμα, αν το δεδομένο είναι ένα βιβλίο, τότε τα μεταδεδομένα του είναι τα στοιχεία που συνθέτουν την περιγραφή του όπως ο συγγραφέας, ο εκδότης, ο τίτλος κα.

Τα μεταδεδομένα διευκολύνουν την εύρεση ενός δεδομένου μέσω ενός δομημένου συνόλου περιγραφών.

Στην ευρύτερη έννοιά τους, τα μεταδεδομένα μπορούν να χρησιμοποιηθούν για να περιγράψουν τις δομές πληροφόρησης όπως τα τεχνικά πρότυπα και τις πολιτικές της διασύνδεσης.

Μπορούμε να πούμε ότι κάθε πληροφορία είναι διακριτή μόνο στο βαθμό που έχει ιδιαίτερα χαρακτηριστικά και «μεταδεδομένα» είναι ο γενικός όρος για τα δεδομένα που χαρακτηρίζουν κάθε είδους πληροφορία. Οι άνθρωποι χαρακτηρίζονται από βιογραφικά μεταδεδομένα, μια τοποθεσία από γεωπληροφορικά μεταδεδομένα,. Στην περίπτωση του ηλεκτρονικού αρχείου, μπορούμε να ορίσουμε μεταδεδομένα, όπως το όνομα, τον ιδιοκτήτη, ημερομηνία δημιουργίας, και τη μορφή. Τα μεταδεδομένα μπορούν να ενσωματωθούν μέσα στο αρχείο, όπως σε ένα έγγραφο επεξεργασίας κειμένου ή μια ιστοσελίδα με ενσωματωμένο τίτλο. Τα μεταδεδομένα μπορούν επίσης να διατηρούνται ξεχωριστά, όπως ένα αρχείο καταλόγου της βιβλιοθήκης είναι ξεχωριστό από ένα βιβλίο. Σε κάθε περίπτωση, η ευκολία στην εύρεση του δεδομένου που ενδιαφέρει τον τελικό χρήστη είναι δεδομένη και εύλογη.

Όσον αφορά στα βιβλία, ενώ υπάρχουν πολλές εκατοντάδες διαφορετικά χαρακτηριστικά που χρησιμοποιούνται για την βιβλιογραφική καταγραφή και την καταγραφή κλάδων με τους οποίους αυτά σχετίζονται, μόνο μερικά χαρακτηριστικά συνήθως χρησιμοποιούνται για την εύρεση ενός βιβλίου, όπως, Τίτλος, Συγγραφέας, Θέμα, Ημερομηνία και Τόπος. Δεν αποτελεί έκπληξη, αυτά τα λίγα χαρακτηριστικά που παραδοσιακά χρησιμοποιούνται για αναφορές σε βιβλία να είναι τα ίδια χαρακτηριστικά που τώρα χρησιμοποιούνται πιο συχνά ως μεταδεδομένα για τα ηλεκτρονικά αρχεία που δημοσιεύονται στον Παγκόσμιο Ιστό.

Διαφαίνεται η ανάγκη ορισμού συγκεκριμένης δομής των μεταδεδομένων καθώς ο υπολογιστής δεν είναι ευέλικτος όσο ο άνθρωπος κατά τη διαδικασία της αναζήτησης δεδομένων. Ποιος ορίζει τη δομή αυτή;

2.5.1 Πρότυπα Μεταδεδομένων

Η προτυποποίηση των διαδικασιών και πολιτικών διαχείρισης εγγραφών διασφαλίζει την απαιτούμενη προσοχή και προστασία στις εγγραφές δεδομένων και την εξόρυξη της πληροφορίας που περιέχουν, χρησιμοποιώντας προτυποποιημένες πρακτικές και διαδικασίες. Το πρότυπο παρέχει οδηγίες στη διαχείριση των εγγραφών δεδομένων των δημόσιων ή ιδιωτικών οργανισμών για εσωτερικούς ή εξωτερικούς πελάτες.

Το πρότυπο εφαρμόζεται στη διαχείριση εγγραφών όλων των μορφοποιήσεων ή των μέσων μετάδοσης που δημιουργούνται ή παραλαμβάνονται από οποιοδήποτε δημόσιο ή ιδιωτικό οργανισμό και παρέχει οδηγίες για τον προσδιορισμό των ευθυνών των αντίστοιχων οργανισμών όσο αφορά στα δεδομένα και τη διαχείριση τους.

Παρέχει τεχνικές και βασικές οδηγίες για αποτελεσματική συγκρότηση, ανάπτυξη και διατήρηση των διοικητικών αρχείων των δημόσιων υπηρεσιών συνεπώς εξασφαλίζει ομοιόμορφες και κοινές αρχές και τεχνικές περιγραφής των διοικητικών εγγράφων

Τα πρότυπα μεταδεδομένων ορίζουν τις απαιτούμενες προδιαγραφές ώστε να εδραιωθεί μια κοινή αντίληψη και κατανόηση του σημασιολογικού περιεχομένου των δεδομένων και να διαβεβαιωθεί η ορθή χρήση και διαλειτουργικότητά τους από τους ιδιοκτήτες και τους τελικούς χρήστες των δεδομένων αυτών των δεδομένων. Αποτελούν πρότυπα σχήματα μεταδεδομένων που αναπτύχθηκαν και συντηρούνται από διεθνείς οργανισμούς προτυποποίησης (όπως ο οργανισμός ISO) ή οργανισμούς που έχουν αναλάβει παρόμοιες πρωτοβουλίες (όπως η

πρωτοβουλία του Dublin Core για τα μεταδεδομένα). Έχουν αναπτυχθεί πολλά και ποικίλα τέτοια πρότυπα[Υποσημείωση] και η διαφορετικότητά τους οφείλεται στον τομέα δεδομένων στον οποίο αναφέρονται όπως βιβλιοθηκονομία, εκπαίδευση, ηλεκτρονική διακυβέρνηση, τέχνες και πολιτισμός κα. Στον πίνακα 1.1 γίνεται αναφορά στα πρότυπα μεταδεδομένων που είναι διαθέσιμα για παρούσα χρήση. Διατηρήθηκε η αγγλική γλώσσα όπως παρουσιάζονται στην αντίστοιχη πηγή αναφοράς τους:

Πίνακας 1.1. Πρότυπα μεταδεδομένων		
Όνομασία προτύπου - πηγή αναφοράς	Τομέας	Περιγραφή
CERIF Common European Research Information Format http://www.eurocris.org	Research	Common European Research Information Format - IS an EU recommendation to member states used widely as the national system for research information in 8 countries and in many organisations in other countries
COSATI Committee on Scientific and Technical Information http://www.dtic.mil/dtic/pdf/submit/Catguide.pdf	Scientific and Technical information	Committee on Scientific and Technical Information - cataloging rules and record format for the descriptive cataloging of technical reports and similar documents.
CSDGM Content Standard for Digital Geospatial Metadata http://www.fgdc.gov/metadata/csdgm/	Geographic data	Content Standard for Digital Geospatial Metadata maintained by the Federal Geographic Data Committee (FGDC).
CSMD Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html	Scientific data sets	Core Scientific Metadata Model - a complex, structured metadata environment for describing research datasets and their context. The standard commonly referred to as FGDC (although FGDC is the maintenance agency, and CSDGM is the actual element set) is a large and early metadata standard for geospatial information created by agencies of the US federal government. The FGDC web site describes the scope of this standard as to allow users to “determine the availability of a set of geospatial data, to determine the fitness [of] the set of geospatial data for an intended use, to determine the means of accessing the set of geospatial data, and to successfully transfer the set of geospatial data.” The current production version of FGDC is 2.0, from 1998. Since this time, an international standard for geospatial information (ISO 19115) has emerged. Plans have been announced to create a US national geospatial metadata standard as a profile of ISO 19115, and to create version 3.0 of CSDGM as an implementation of that.
DC Dublin Core Metadata Element Set http://www.dublincore.org/documents/dces/	Network sources	Dublin Core - Interoperable online metadata standard focused on networked resources, with worldwide and cross-domain acceptance.

2.6 Πηγές Ανοικτών Δεδομένων

Η ηλεκτρονική διακυβέρνηση είναι υπαρκτή στην Ελλάδα περισσότερο από μία δεκαετία και επεκτείνεται καθώς οι περισσότεροι φορείς της Δημόσιας Διοίκησης, διαθέτουν, αναπτύσσουν ή σχεδιάζουν να αναπτύξουν πληροφοριακά συστήματα με σκοπό να παρέχουν πληροφορίες και υπηρεσίες σε πολίτες, επιχειρήσεις και άλλους φορείς ^[25].

Ως πηγές γνώσεις του Δημόσιου Τομέα θεωρούμε όλους τους φορείς του στενού και ευρύτερου τομέα που προσφέρουν ανοιχτή πρόσβαση σε κυβερνητικά δεδομένα. Αναλυτικότερα, ο Πρωθυπουργός, το Υπουργικό Συμβούλιο και τα συλλογικά κυβερνητικά όργανα, οι Υπουργοί, Αναπληρωτές Υπουργοί, Υφυπουργοί, Γενικοί Γραμματείς Υπουργείων και Περιφερειών, Ειδικοί Γραμματείς Υπουργείων, κάθε κεντρική, ειδική ή περιφερειακή δημόσια υπηρεσία, τα όργανα διοίκησης νομικών προσώπων δημοσίου δικαίου (Ν.Π.Δ.Δ.), οι ανεξάρτητες και ρυθμιστικές διοικητικές αρχές, το Νομικό Συμβούλιο του Κράτους, τα όργανα διοίκησης φορέων του ευρύτερου δημόσιου τομέα καθώς και τα όργανα των φορέων των Οργανισμών Τοπικής Αυτοδιοίκησης πρώτου και δεύτερου βαθμού. Επίσης όργανα στα οποία οι προηγούμενοι έχουν χορηγήσει εξουσιοδότηση προς υπογραφή ή έχουν μεταβιβάσει αρμοδιότητα.

Επιπρόσθετα, οργανισμοί του ευρύτερου Δημόσιου και του Ιδιωτικού Τομέα, οι οποίοι διαδραματίζουν σημαντικό ρόλο στην ανάπτυξη και παροχή ηλεκτρονικών υπηρεσιών προς πολίτες, επιχειρήσεις και άλλους φορείς, και αναλυτικότερα:

- Δημόσιες Επιχειρήσεις Κοινής Ωφέλειας
- Τραπεζικούς και Χρηματοπιστωτικούς Οργανισμούς
- Επιχειρήσεις Πληροφορικής και Υπηρεσιών που δραστηριοποιούνται στην ανάπτυξη λογισμικού και την παροχή συναφών υπηρεσιών για φορείς της Δημόσιας Διοίκησης (π.χ. ανάπτυξη διαδικτυακών πυλών, υλοποίηση πληροφοριακών συστημάτων).

Συγκεντρωτικά, τα ανοιχτά δημόσια δεδομένα στην Ελλάδα συνοψίζονται στον παρακάτω πίνακα:

Πίνακας 1.2. Πηγές ανοικτών δημόσιων δεδομένων στην Ελλάδα

A/A	Φορέας	Περιγραφή	Είδος πληροφοριών	Ηλεκτρονική διεύθυνση
1	Πρόγραμμα Διαύγεια	Νόμοι / Κανονισμοί / Αποφάσεις Υπουργείων & Δημόσιες Δαπάνες Φορέων	Οικονομικά / Διακυβέρνηση	http://diavgeia.gov.gr http://greek-lod.math.auth.gr/diavgeia/
2	Γενική Γραμματεία Πληροφοριακών Συστημάτων	Στατιστικά Στοιχεία Φόρων Της Ελλάδας	Οικονομικά	http://www.gsis.gr
3	Geodata	Ένα Έργο Που Χρηματοδοτείται Από Την Ελληνική Κυβέρνηση, Προκειμένου Να Φιλοξενήσει Γεωχωρικά Δημόσια Δεδομένα	Περιβάλλον, Δημογραφικά, Γεωχωρικά, Εγκληματικότητα	http://geodata.gov.gr
4	Ελληνική Στατιστική Αρχή	Η Επίσημη Πηγή Για Τις Στατιστικές Πληροφορίες Για Σχεδόν Κάθε Δραστηριότητα Της Ελληνικής Κυβέρνησης	Στατιστικά	http://www.statistics.gr
5	Υπουργείο Οικονομικών	Οικονομικά Στοιχεία Της Χώρας	Οικονομικά, Νόμοι	http://www.minfin.gr
6	Πυροσβεστικό Σωμα Ελλαδος	Παροχή Πληροφοριών Σχετικά Με Τις Πυρκαγιές Που Έχουν Καταγραφεί Στην Ελλάδα	Περιβάλλον, Γεωχωρικά	http://www.fireservice.gr http://greek-lod.math.auth.gr/fire-brigade/
7	Υπουργείο Εργασίας, Κοινωνικής Ασφάλισης & Πρόνοιας	Παροχή Νομικών Πληροφοριών Σχετικά Με Τα Δικαιώματα Απασχόλησης Και Τις Αποφάσεις Του Υπουργείου	Νόμοι	http://www.ypakp.gr
8	Υπουργείο Παιδείας & Θρησκευμάτων, Πολιτισμού & Αθλητισμού	Παροχή Πληροφοριών Από Την Τοποθεσία Όλων Των Σχολείων Στην Ελλάδα	Γεωχωρικά / Γραφικά	http://www.minedu.gov.gr/

9	Ελληνική Αστυνομία	Παροχή Πληροφοριών Για Τα Τροχαία Ατυχήματα Και Την Εγκληματικότητα	Στατιστικά / Γεωχημικά	http://www.astynomia.gr http://greek-lod.math.auth.gr/police/
11	Υπουργείο Υποδομών Μεταφορών & Δικτύων	Λεπτομέρειες Για Τιμολόγια Δημοσίων Έργων	Οικονομικά / Νόμοι	http://www.yme.gr/
12	Υπουργείο Υγείας & Κοινωνικής Αλληλεγγύης	Παρέχει Συγκεντρωτικές Πληροφορίες Σχετικά Με Χειρουργικές Επεμβάσεις Και Τις Δοκιμές Που Διεξάγονται Σε Νοσοκομεία	Στατιστικά / Υγεία	http://yyka.gov.gr
13	Υπουργείο Αγροτικής Ανάπτυξης & Τροφίμων	Παροχή Επιχορηγήσεων Ανά Νομό Και Χρηματοδοτικές Υποθέσεις	Στατιστικά / Εικονικά	http://www.minagric.gr
14	Υπουργείο Περιβάλλοντος Ενέργειας & Κλιματικής Αλλαγής	Πληροφορίες Σχετικά Με Τις Προστατευόμενες Περιοχές, Όπως Ζώνες Προστασίας Βουνού, Εθνικούς Δρυμούς Και Δάση, Ελεγχόμενες Περιοχές Κυνηγιού, Κλπ.	Περιβάλλον / Γεωχωρικά	http://www.ypeka.gr
15	Παρατηρητήριο Για Την Ψηφιακή Ελλάδα	Παρέχει Πληροφορίες Σχετικά Με Τους Δείκτες Της Κοινωνίας Της Πληροφορίας	Στατιστικά	http://www.observatory.gr/
16	Ελλάνικο Lod Cloud	Πρώτη Προσπάθεια Για Την Παροχή Διασυνδεδεμένων Ανοιχτών Δεδομένων Όχι Μόνο Κυβερνητικών Δεδομένων (Αστυνομία, Πυροσβεστική, Διαύγεια, Καλλικράτης)	Απογραφή, Γεωχωρικά, Οικονομικά, Κυβερνητικά	http://greek-lod.math.auth.gr/
17	Οργανισμός Αστικών Συγκοινωνιών Αθηνών	Δημόσιες Μεταφορές Και Τις Στάσεις	Μεταφορές, Οικονομικά	http://www.oasa.gr

18	Αρχαιολογικό Κτηματολόγιο	Ποσοτικά Και Γεωχωρικά Δεδομένα, Για Αρχαιολογικούς Χώρους	Γεωχωρικά, Στατιστικά	http://archaeocadastre.culture.gr/el/data
19	Γενική Γραμματεία Αιγαίου Και Νησιώτικης Πολίτικης	Παροχή Πληροφοριών Σχετικά Με Οποιοδήποτε Θέμα Που Λαμβάνει Χώρα Στα Νησιά Του Αιγαίου	Περιβάλλον	http://www.ypai.gr/
20	Υπουργείο Αναπτύξεις, Ανταγωνιστικότητας & Ναυτιλίας	Παροχή Πληροφοριών Σχετικά Με Τα Δημόσια Έργα	Κυβερνητικά, Οικονομικά, Περιβάλλον	http://www.ypoian.gr/
21	Γενική Γραμματεία Έρευνας & Τεχνολογίας	Παροχή Πληροφοριών Σχετικά Με Τα Δημόσια Έργα	Κυβερνητικά, Οικονομικά, Περιβάλλον	http://www.gsrt.gr/
22	Υπουργείο Εθνικής Άμυνας	Παροχή Πληροφοριών Σχετικά Με Ελληνικό Στρατό	Στατιστικά, Οικονομικά	http://www.mod.gr/
23	Υπουργείο Δημόσιας Τάξης & Προστασίας Του Πολίτη	Παροχή Πληροφοριών Για Την Ελληνική Αστυνομία, Πυροσβεστική Υπηρεσία, Αστυνομία Κλπ.	Στατιστικά, Περιβάλλον, Κυβερνητικά	http://www.minocp.gov.gr/
25	Υπουργείο Θαλάσσιων Υποθέσεων, Νήσων & Αλιείας	Παροχή Πληροφοριών Για Τα Ελληνικά Νησιά, Την Αλιεία Και Τις Θάλασσες	Περιβάλλον, Κυβερνητικά	http://www.yen.gr
26	Υπουργείο Εξωτερικών	Παροχή Πληροφοριών Σχετικά Με Την Εξωτερική Πολιτική Της Ελλάδος	Κυβερνητικά	http://www.mfa.gr/
27	Υπουργείο Εσωτερικών	Παροχή Πληροφοριών Σχετικά Με Την Εσωτερική Πολιτική Της Ελλάδος	Κυβερνητικά, Οικονομικά	http://www.ypes.gr/

28	Υπουργείο Μακεδονίας Θράκης	- Παροχή Πληροφοριών Σχετικά Με Κάθε Θέμα Για Την Βόρεια Ελλάδα	Κυβερνητικά, Οικονομικά	http://www.mathra.gr/
29	Υπουργείο Πολιτισμού Και Τουρισμού	Παροχή Πληροφοριών Για Τον Ελληνικό Πολιτισμό Και Τον Τουρισμό	Κυβερνητικά, Οικονομικά Μεταφορές	http://www.culture.gr
30	Γενική Γραμματεία Ενημέρωσης & Επικοινωνίας	Παροχή Πληροφοριών Σχετικά Με Τα ΜΜΕ	Κυβερνητικά, Οικονομικά	http://www.minpress.gr/
31	Νομαρχεία Αθηνών	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στην Αθήνα	Κυβερνητικά, Οικονομικά	http://nom-athinas.gov.gr
32	Περιφέρεια Ανατολικής Μακεδονίας Θράκης	- Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στην Μακεδονία - Θράκη	Κυβερνητικά, Οικονομικά	http://www.pamth.gov.gr/
33	Περιφέρεια Νοτίου Αιγαίου	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στο Νότιο Αιγαίο	Κυβερνητικά, Οικονομικά	http://www.notioaigaio.gr/per/
34	Αποκεντρωμένη Διοίκηση Ηπείρου – Δυτικής Μακεδονίας	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στην Ήπειρο – Δυτική Μακεδονία	Κυβερνητικά, Οικονομικά	http://www.epirus.gov.gr
35	Αποκεντρωμένη Διοίκηση Θεσσαλίας & Στερεάς Ελλάδος	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Θεσσαλονίκη	Κυβερνητικά, Οικονομικά	http://www.thessalia.gov.gr/
36	Περιφέρεια Ιονίων Νήσων	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στα Ιόνια Νησιά	Κυβερνητικά, Οικονομικά	http://www.pin.gov.gr/
37	Περιφέρεια δυτικής Ελλάδας	Παροχή πληροφοριών σχετικά με θέματα που λαμβάνουν χώρα στη δυτική Ελλάδα	Κυβερνητικά, Οικονομικά	http://www.pde.gov.gr/gr/index.php

38	Αποκεντρωμένη διοίκηση αττικής	Παροχή πληροφοριών σχετικά με θέματα που λαμβάνουν χώρα στην αττική	Κυβερνητικά, Οικονομικά	http://www.attiki.gov.gr/
39	Περιφέρεια Πελοποννήσου	Παροχή πληροφοριών σχετικά με θέματα που λαμβάνουν χώρα στη Πελοπόννησο	Κυβερνητικά, Οικονομικά	http://ppel.gov.gr/
40	Περιφέρεια βορείου αιγαίου	Παροχή πληροφοριών σχετικά με θέματα που λαμβάνουν χώρα στο βόριο αιγαίο	Κυβερνητικά, Οικονομικά	http://www.northaegaeon.gr/
41	Περιφέρεια κεντρικής Μακεδονίας	Παροχή πληροφοριών σχετικά με θέματα που λαμβάνουν χώρα στη κεντρική Μακεδονία	Κυβερνητικά, Οικονομικά	http://www.pkm.gov.gr/
42	Δήμος Κερατσινίου Δραπετσώνας	Παροχή πληροφοριών σχετικά με θέματα που λαμβάνουν χώρα στο Κερατσίνι - Δραπετσώνα	Κυβερνητικά, Οικονομικά	http://www.keratsinidrapetsona.gr/
43	Δήμος αθηναίων	Παροχή πληροφοριών σχετικά με θέματα που λαμβάνουν χώρα στην Αθήνα	Κυβερνητικά, Οικονομικά	http://www.cityofathens.gr/
44	Δήμος Θεσσαλονίκης	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Θεσσαλονίκη	Κυβερνητικά, Οικονομικά	http://www.thessaloniki.gr
45	Δήμος Χανίων	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στα Χανιά	Κυβερνητικά, Οικονομικά	http://www.chania.gr/
46	Δήμος Ρέθυμνου	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στο Ρέθυμνο	Κυβερνητικά, Οικονομικά	http://www.rethymno.gr/
47	Δήμος Ηρακλείου	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στο Ηράκλειο	Κυβερνητικά, Οικονομικά	http://www.heraklion.gr/

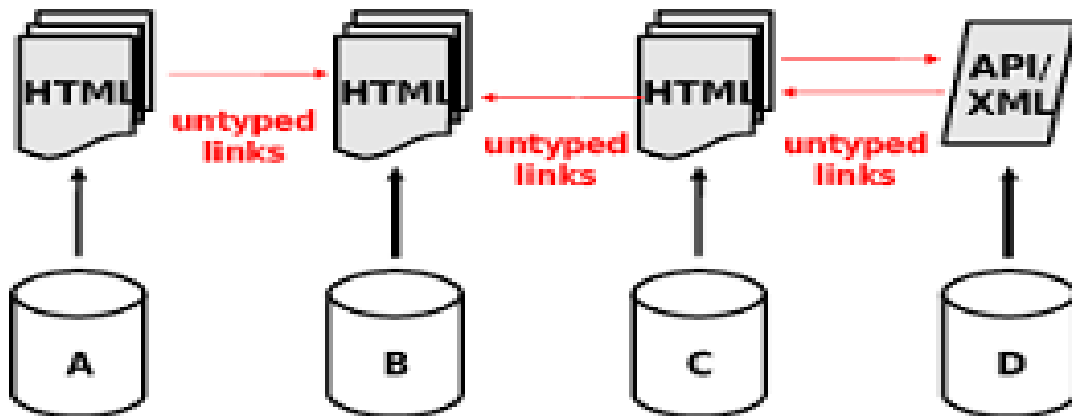
48	Δήμος Λαρίσης	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Λάρισα	Κυβερνητικά, Οικονομικά	http://www.larissa-dimos.gr/
49	Δήμος Μεσσηνίας	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Μεσσηνία	Κυβερνητικά, Οικονομικά	http://www.messini.gr
50	Δήμος Σάμου	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Σάμο	Κυβερνητικά, Οικονομικά	http://www.vathi.gr/
51	Δήμος Χίου	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Χίο	Κυβερνητικά, Οικονομικά	http://www.chioscity.gr/
52	Δήμος Μυτιλήνης	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Μυτιλήνη	Κυβερνητικά, Οικονομικά	http://www.mytilene.gr/
53	Δήμος Χαλκιδικής	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Χαλκιδική	Κυβερνητικά, Οικονομικά	http://www.halkidiki.gov.gr
54	Δήμος Δράμας	Παροχή Πληροφοριών Σχετικά Με Θέματα Που Λαμβάνουν Χώρα Στη Δράμα	Κυβερνητικά, Οικονομικά	http://www.dimos-dramas.gr

ΚΕΦΑΛΑΙΟ 3 - Σημασιολογικός Ιστός, Συνδεδεμένα Δεδομένα και Συνδεδεμένα Κυβερνητικά Δεδομένα

3.1 Από τον Ιστό των Εγγράφων στον Ιστό των Δεδομένων

Η ιδέα του **Παγκόσμιου Ιστού (World Wide Web)** ανήκει στον **Tim Berners-Lee**. Το 1989 και ενώ εργαζόταν στα εργαστήρια του CERN της Γενεύης, είχε την ιδέα να δημιουργήσει ένα σύστημα που θα επέτρεπε σε ερευνητές από απομακρυσμένα μεταξύ τους μέρη του κόσμου να οργανώσουν και να συνδέσουν όλο τον όγκο των εγγράφων που διέθεταν. Πρότεινε την εισαγωγή συνδέσμων (links) μέσα στο κείμενο του κάθε εγγράφου ώστε να είναι δυνατή η πρόσβαση από οποιονδήποτε σε οποιοδήποτε έγγραφο. Έτσι, καθώς κάποιος διαβάζει ένα έγγραφο θα μπορεί μέσω των links να δει και να διαβάσει όλα τα σχετικά, όλα τα συνδεδεμένα με το αρχικό, έγγραφα. Ως αποτέλεσμα, ο Παγκόσμιος Ιστός, όπως τον ξέρουμε μέχρι σήμερα, αποτελείται από έγγραφα (που αποκαλούνται Web documents ή Web pages ή HTML pages) τα οποία συνδέονται μεταξύ τους με **υπερσυνδέσμους (hyperlinks)**. Αναγκαία ήταν η δημιουργία μιας γλώσσας κατανοητής από όλους τους διαφορετικούς υπολογιστές και παράλληλα ενός τρόπου επικοινωνίας, ενός πρωτοκόλλου δηλαδή μεταξύ των υπολογιστών, ώστε να είναι δυνατή η ανάκτηση, το «κατέβασμα» κοινώς, των εγγράφων με την βοήθεια των συνδέσμων. Έτσι δημιουργήθηκε η γλώσσα οδηγιών HTML (Hypertext Markup Language) και το πρωτόκολλο HTTP (Hypertext Transfer Protocol).^[1]

Ο Παγκόσμιος Ιστός είναι ανάλογος με ένα παγκόσμιο σύστημα αρχείων, με πρωταρχικά αντικείμενα τα έγγραφα και τους συνδέσμους μεταξύ αυτών. Τα αντικείμενα αυτά είναι δομημένα σε πολύ μικρό βαθμό ενώ η σημασιολογία (semantics) του περιεχομένου και των συνδέσμων δεν είναι φανερή. Ο Ιστός είναι κυρίως σχεδιασμένος για ανθρώπινη κατανάλωση. Για τους λόγους αυτούς, ο Παγκόσμιος Ιστός λέγεται και Ιστός των Εγγράφων (Web of Documents).^[2]



Σημα 3.1 : Ιστός Εγγράφων [2]

Στο σημείο αυτό να επισημάνουμε ότι η αρχική ιδέα του Ιστού όπως το είχε οραματιστεί ο Tim Berners-Lee περιελάμβανε τόσο έγγραφα όσο και δεδομένα [18] (καθώς και πιο αφηρημένες έννοιες). Υποθέτουμε ότι ένας από τους κύριους λόγους για τους οποίους το κομμάτι των δεδομένων δεν τα κατάφερε σε πρώτη φάση να εξαπλωθεί, ίσως να ήταν ότι οι άνθρωποι και τα ιδρύματα επικεντρώθηκαν στο κομμάτι της παρουσίασης (όπως η διάταξη, η διάδραση και η δομή) παρά στα πρωτογενή δεδομένα. Παρά τα αδιαμφισβήτητα οφέλη που παρέχει ο Ιστός, μέχρι πρόσφατα οι ίδιοι κανόνες που έδωσαν στον Ιστό των Εγγράφων τη δυνατότητα να ανθίσει δεν εφαρμόζονταν στα δεδομένα με αποτέλεσμα ο Ιστός να είναι κάπως ατελής. Τα πρώτα χρόνια (μιλώντας κατά προσέγγιση από τις αρχές του '90 έως το 2000), οι περισσότεροι άνθρωποι ήταν απασχολημένοι με την ανάπτυξη λογισμικού όπως πράκτορες και εξυπηρετητές, αλλά και περιεχομένου για τον Ιστό των Εγγράφων. [4]

Σύμφωνα με τον Tim Berners-Lee, ο Ιστός σήμερα είναι αρκετά αποτελεσματικός στο να μας βοηθά να δημοσιεύουμε και να ανακαλύπτουμε έγγραφα, αλλά τα μεμονωμένα στοιχεία πληροφοριών μέσα σε αυτά τα έγγραφα (είτε είναι η ημερομηνία ενός γεγονότος, η τιμή ενός αντικειμένου σε ένα κατάλογο ή ένας μαθηματικός τύπος) δεν μπορούν να υποστούν επεξεργασία απ' ευθείας ως δεδομένα. Σήμερα μπορούμε να δούμε τα δεδομένα με ένα φυλλομετρητή (browser), αλλά δεν μπορούμε να χρησιμοποιήσουμε άλλα προγράμματα για το χειρισμό ή την ανάλυσή τους χωρίς πρώτα να τα έχουμε επεξεργαστεί εμείς οι ίδιοι χειρωνακτικά. Αν λυθεί το πρόβλημα αυτό, ο Ιστός θα αναμένεται να μοιάζει περισσότερο με μεγάλη βάση δεδομένων ή φύλλο εργασίας (spreadsheet), παρά ως ένα απλό σύνολο συνδεδεμένων εγγράφων. Οι ψηφιακές πληροφορίες που αφορούν σχεδόν κάθε πλευρά της ζωής μας

δημιουργούνται με ραγδαίους ρυθμούς. Μέσα σε όλα αυτά τα δεδομένα βρίσκεται το κλειδί της γνώσης για τη θεραπεία ασθενειών, τη δημιουργία επιχειρηματικής αξίας και της πιο αποτελεσματικής διακυβέρνησης του κόσμου μας. [3] Αν και ο Παγκόσμιος Ιστός είναι η μεγαλύτερη αποθήκη πληροφοριών που δημιουργήθηκε ποτέ, με τα περιεχόμενά του να επεκτείνονται σε διάφορες γλώσσες και πεδία γνώσεων, μακροπρόθεσμα, είναι εξαιρετικά δύσκολο να βγει νόημα από το περιεχόμενό του. Οι μηχανές αναζήτησης μπορεί να βοηθούν στην εύρεση περιεχομένου που περιέχει συγκεκριμένες λέξεις, αλλά το περιεχόμενο μπορεί να μην είναι ακριβώς αυτό που θέλουμε. Η αναζήτηση βασίζεται στα περιεχόμενα των σελίδων και όχι στο εννοιολογικό νόημα των περιεχομένων της σελίδας ή των πληροφοριών για τη σελίδα. ^[4]

Από το 2000 έως το 2006 (χονδρικά) ένας αριθμός τεχνικών καινοτομιών, το RDF το οποίο είναι για τα δεδομένα ό,τι η HTML για τα έγγραφα και η Web Ontology Language (OWL) η οποία μας επιτρέπει να εκφράσουμε το πώς συνδέονται μεταξύ τους οι πηγές δεδομένων, σε συνδυασμό με τις πιο ανοιχτές πρακτικές διαμοιρασμού πληροφοριών, μετακινούν τον Παγκόσμιο Ιστό προς αυτό που ονομάζουμε **Σημασιολογικό Ιστό**. Η ύπαρξη του Σημασιολογικού Ιστού δίνει τη δυνατότητα να τοποθετηθούν ετικέτες σε όλο το περιεχόμενο του Ιστού, να περιγράψουμε τι είναι κάθε κομμάτι πληροφορίας και να δώσουμε σημασιολογικό νόημα στο αντικείμενο του περιεχομένου. Έτσι, οι μηχανές αναζήτησης θα γίνουν πιο αποτελεσματικές απ' ό,τι είναι τώρα και οι χρήστες θα μπορούν να βρουν τις ακριβείς πληροφορίες που ψάχνουν. Οι οργανισμοί που παρέχουν διάφορες υπηρεσίες μπορούν να βάλουν ετικέτες με νόημα σε αυτές τις υπηρεσίες. Χρησιμοποιώντας πράκτορες λογισμικού βασισμένους στον Ιστό, μπορούμε να βρούμε δυναμικά αυτές τις υπηρεσίες και να τις χρησιμοποιήσουμε προς όφελός μας ή σε συνεργασία με άλλες υπηρεσίες.

	Ιστός Εγγράφων	Σημασιολογικός Ιστός
Βασική οντότητα	Έγγραφο (html document)	Αντικείμενα (ή περιγραφές αντικειμένων)
Διασυνδέσεις μεταξύ	Εγγράφων	Αντικειμένων
Βαθμός δομής κάθε οντότητας	Χαμηλή	Υψηλή
Σημασία του περιεχομένου και των συνδέσμων	Υποκειμενική	Αντικειμενική
Σχεδιασμένο για κατανόηση	Πρωταρχικά από ανθρώπους, δευτερευόντως από μηχανές	Πρωταρχικά από μηχανές, δευτερευόντως από ανθρώπους

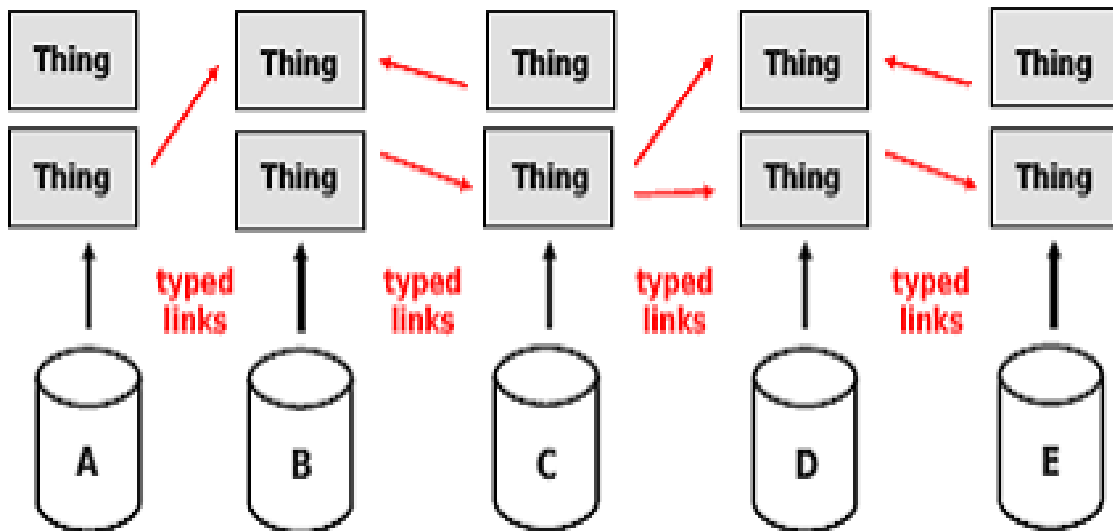
Σχήμα 3.2 : Διαφορές μεταξύ Σημασιολογικού Ιστού και Ιστού Δεδομένων

Η πρόοδος προς την καλύτερη ολοκλήρωση των δεδομένων θα συντελεστεί μέσω της χρήσης της τεχνολογίας-κλειδί που έκανε τον Παγκόσμιο Ιστό τόσο επιτυχημένο: το σύνδεσμο. Η δύναμη του Ιστού σήμερα, συμπεριλαμβανομένης της δυνατότητας να βρούμε τη σελίδα για την οποία ψάχνουμε, έγκειται στο γεγονός τα έγγραφα τοποθετούνται στον Ιστό σε πρότυπη μορφή και έπειτα συνδέονται μεταξύ τους. Ο Σημασιολογικός Ιστός θα δώσει τη δυνατότητα καλύτερης ολοκλήρωσης των δεδομένων, επιτρέποντας στον καθένα ο οποίος τοποθετεί μεμονωμένα στοιχεία δεδομένων στον Ιστό, να τα συνδέσει με άλλα κομμάτια δεδομένων, χρησιμοποιώντας πρότυπες μορφοποιήσεις.

Το αποτέλεσμα της χρήσης των τεχνολογιών του Σημασιολογικού Ιστού για να:

1. δημοσιεύσουμε δομημένα δεδομένα στον Ιστό,
2. θέσουμε συνδέσμους μεταξύ δεδομένων από μια πηγή δεδομένων προς δεδομένα μέσα σε άλλες πηγές δεδομένων είναι τα **Συνδεδεμένα Δεδομένα (Linked Data)**.^[3]

Ως αποτέλεσμα θα προκύψει ο Ιστός των Δεδομένων (Web of Data). Με τη διασύνδεση συνόλων δεδομένων, οι μηχανές θα μπορούν να διασχίζουν έναν ανεξάρτητο ιστό δομημένων πληροφοριών έτσι ώστε να συλλέξουν σημασιολογική γνώση από αυθαίρετες οντολογίες και τομείς. Το αποτέλεσμα θα είναι μια μαζική βάση γνώσης με ελεύθερη πρόσβαση η οποία θα αποτελέσει τα θεμέλια μιας νέας γενιάς εφαρμογών και υπηρεσιών.



Σχήμα 3.3 : Ιστός Συνδεδεμένων Δεδομένων

3.2 Κανόνες Συνδεδεμένων Δεδομένων

Ο **Tim Berners-Lee** ήταν ο εμπνευστής της ιδέας του Παγκόσμιου Ιστού αλλά και της ιδέας των Συνδεδεμένων Δεδομένων που θα έχει ως αποτέλεσμα τον Ιστό των Δεδομένων. Το 2006, ο Tim Berners-Lee έγραψε ένα σχεδιαστικό σημείωμα, προτείνοντας βασικά τη εφαρμογή των ίδιων κανόνων που κατέστησαν επιτυχημένο τον Ιστό των Εγγράφων, στα δεδομένα. Οι **τέσσερις κανόνες των Συνδεδεμένων Δεδομένων** στην αρχική τους ανάγνωση έχουν ως εξής:

1. Όλα τα αντικείμενα πρέπει να ταυτοποιούνται με τη χρήση URIs
2. Όλα τα URIs πρέπει να είναι dereferenceable, δηλαδή η χρήση των HTTP URIs να επιτρέπει στα άτομα να ανατρέξουν στο αντικείμενο που ταυτοποιείται από το URI και να λάβουν μια περιγραφή του
3. Όταν κάποιος ανατρέχει σε ένα URI- δηλαδή σε μια RDF ιδιότητα που μεταφράζεται ως υπερσύνδεσμος- οδηγείται σε περισσότερα χρήσιμα δεδομένα (αναφέρεται και ως κανόνας follow-your-nose)
4. Πρέπει να συμπεριλαμβάνονται σύνδεσμοι προς άλλα URIs ώστε να επιτρέπουν την ανακάλυψη περισσότερων δεδομένων.

Σημειώνεται ότι το 2009 ο Tim Berners-Lee τροποποίησε ελαφρώς τους κανόνες, προσθέτοντας ρητά το Resource Description Framework (RDF) στον τρίτο κανόνα, ενώ οι συζητήσεις για το εάν το RDF είναι πράγματι ουσιώδες μέρος των Συνδεδεμένων Δεδομένων συνεχίζονται. Επίσης, σημειώνεται ότι σε αντίθεση με το ολοκληρωμένο όραμα του Σημασιολογικού Ιστού, τα Συνδεδεμένα Δεδομένα αφορούν κυρίως τη δημοσίευση δομημένων δεδομένων σε RDF χρησιμοποιώντας URIs παρά την εστίαση στο οντολογικό επίπεδο. Αυτή η απλοποίηση – ακριβώς όπως ο Ιστός απλοποίησε τις καθιερωμένες ακαδημαϊκές προσεγγίσεις των συστημάτων Υπερκειμένου – μειώνει τα εμπόδια συμμετοχής για τον πάροχο των δεδομένων και επομένως ενισχύει την ευρεία υιοθέτησή του.

3.3 Θέματα Αρχιτεκτονικής Συνδεδεμένων Δεδομένων

Τα Συνδεδεμένα Δεδομένα βασίζονται σε δύο τεχνολογίες που είναι θεμελιώδεις στον Ιστό: τα Ενιαία Αναγνωριστικά Πόρων (Uniform Resource Identifiers-URIs) και το Πρωτόκολλο Μεταφοράς Υπερκειμένου. Ενώ οι Ενιαίοι Εντοπιστές Πόρων (Uniform Resource Locators-URLs) έγιναν γνωστοί ως διευθύνσεις για έγγραφα και άλλες οντότητες τα οποία μπορούν να τοποθετηθούν στο Web, τα Ενιαία Αναγνωριστικά Πόρων αποτελούν έναν πιο γενικό τρόπο για την ταυτοποίηση κάθε οντότητας που υπάρχει στον κόσμο.

Τα URIs και το HTTP υλοποιούνται με μια τεχνολογία πολύ σημαντική στον Ιστό Δεδομένων – το Resource Description Framework (RDF). Ενώ η HTML παρέχει ένα μέσο για τη δόμηση και τη σύνδεση εγγράφων στον Ιστό, το RDF παρέχει ένα γενικό, βασισμένο σε γράφο, μοντέλο δεδομένων με το οποίο θα δομηθούν και θα συνδεθούν τα δεδομένα που περιγράφουν αντικείμενα στον κόσμο. Το μοντέλο αυτό είναι εξαιρετικά απλό από τη μια μεριά αλλά από την άλλη, αυστηρά προσαρμοσμένο στην αρχιτεκτονική του Ιστού.

Το μοντέλο RDF ⁶ κωδικοποιεί τα δεδομένα σε μορφή τριάδων υποκείμενο, κατηγορημα, αντικείμενο. Το υποκείμενο και το αντικείμενο μιας τριάδας είναι και τα δύο URIs που το καθένα ταυτοποιεί ένα πόρο, ή ένα URI και ένα λεκτικό αλφαριθμητικό, αντίστοιχα. Το κατηγορημα

⁶ <http://www.w3.org/RDF/>

καθορίζει το πώς συνδέονται μεταξύ τους το υποκείμενο και το αντικείμενο και επίσης αναπαριστάται από ένα URI. ^[6]

Στο πλαίσιο των Συνδεδεμένων Δεδομένων η χρήση του μοντέλου RDF αποφέρει οφέλη όπως τα εξής:

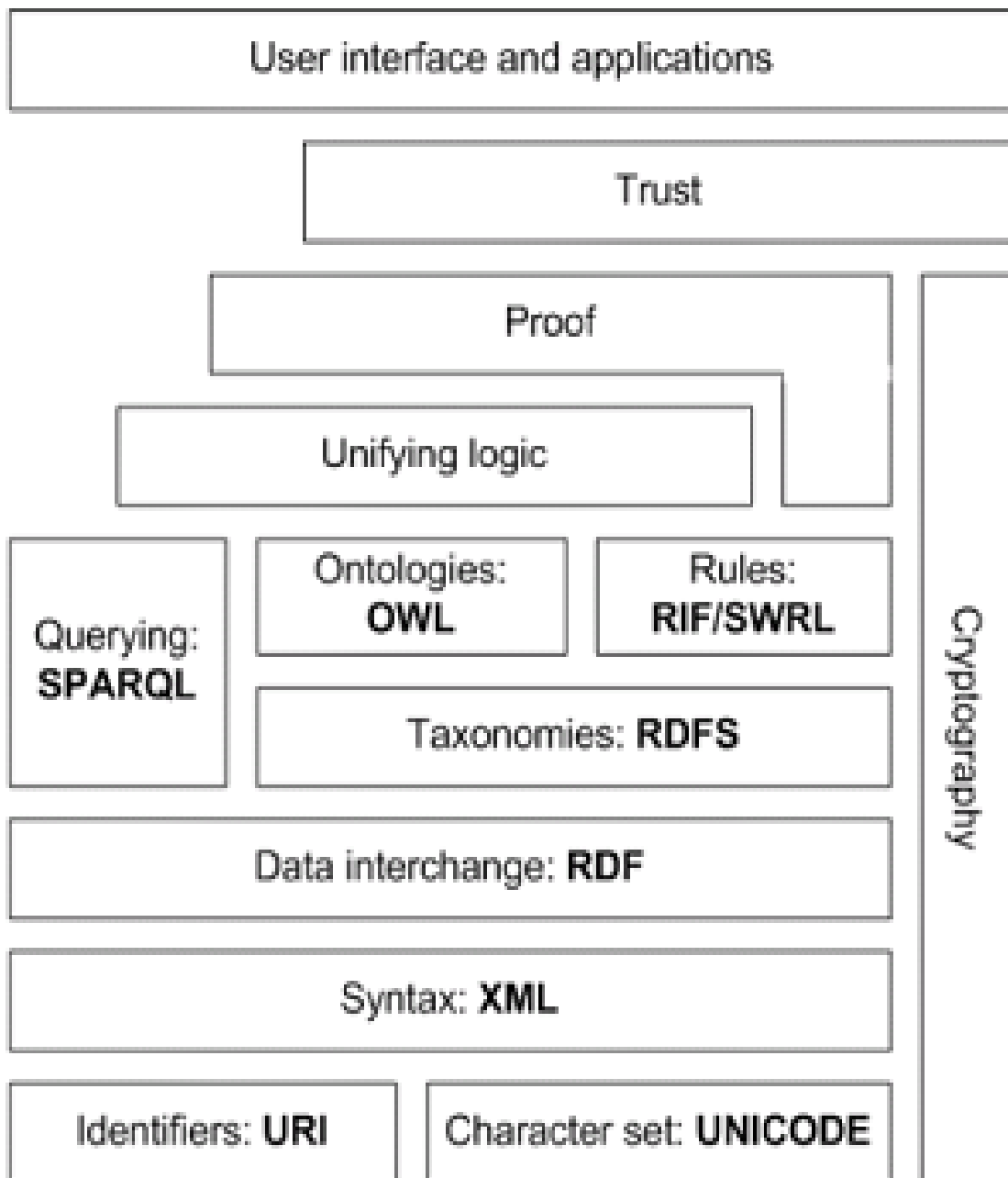
- Οι πελάτες μπορούν να επισκεφθούν κάθε URI ενός RDF γράφου στον Ιστό για να ανακτήσουν επιπλέον πληροφορίες.
- Πληροφορίες από διαφορετικές πηγές συγχωνεύονται με φυσικό τρόπο.
- Το μοντέλο δεδομένων δίνει τη δυνατότητα σε κάποιον να θέσει RDF συνδέσμους μεταξύ δεδομένων από διαφορετικές πηγές.
- Το μοντέλο δεδομένων επιτρέπει σε κάποιον να αναπαραστήσει πληροφορίες που εκφράζονται με τη χρήση διαφορετικών σχημάτων σε ένα μοντέλο.
- Συνδυαζόμενο με σχηματικές γλώσσες όπως η RDF-S και η OWL, το μοντέλο δεδομένων επιτρέπει τη χρήση δομής στο βαθμό που απαιτείται κάθε φορά, που σημαίνει ότι μπορούμε να αναπαραστήσουμε αυστηρά δομημένα δεδομένα καθώς και ημι-δομημένα δεδομένα. ^[7]

Ένα συνηθισμένο σχήμα σύνταξης για τα Συνδεδεμένα Δεδομένα είναι το RDF/XML⁷. Επίσης, χρησιμοποιούνται ως εναλλακτικά συντακτικά η N3⁸ και το υποσύνολό της Turtle⁹. ^[6]

⁷ <http://www.w3.org/TR/REC-rdf-syntax/>

⁸ <http://www.w3.org/2007/02/turtle/primer/>

⁹ <http://www.w3.org/DesignIssues/Notation3.html>



Σχήμα 3.4 : Η Αρχιτεκτονική του Σημασιολογικού Ιστού

3.3.1 Πληροφοριακοί και μη-Πληροφοριακοί Πόροι

Πριν τη δημοσίευση δεδομένων στον Ιστό, το πρώτο βήμα πρέπει να είναι η κατανόηση των δεδομένων μας να βρούμε δηλαδή ποιες είναι οι οντότητες κλειδιά στο σύνολο δεδομένων μας, τι ιδιότητες έχουν και πώς συνδέονται με άλλες οντότητες. Στην ορολογία της Αρχιτεκτονικής του Ιστού, όλα τα αντικείμενα του ενδιαφέροντός μας καλούνται πόροι.

Το W3C Technical Architecture Group (TAG) ξεχωρίζει δύο είδη πόρων: **πληροφοριακοί πόροι και μη-πληροφοριακοί πόροι** (λέγονται και «άλλοι πόροι»). Αυτός ο διαχωρισμός είναι πολύ σημαντικός στο πλαίσιο των Συνδεδεμένων Δεδομένων. Όλοι οι πόροι που βρίσκουμε στον παραδοσιακό Ιστό των Εγγράφων, όπως έγγραφα, εικόνες, και άλλα αρχεία (πολυ)μέσων, είναι πληροφοριακοί πόροι. Αλλά πολλά από τα αντικείμενα για τα οποία θέλουμε να μοιραστούμε δεδομένα δεν είναι: άνθρωποι, απτά αντικείμενα, τοποθεσίες, πρωτεΐνες, επιστημονικές ιδέες κ.ο.κ. Πρωταρχικός κανόνας είναι ότι όλα τα «αντικείμενα του πραγματικού κόσμου» τα οποία υπάρχουν έξω από τον Ιστό είναι μη-πληροφοριακοί πόροι.

Οι πληροφοριακοί πόροι μπορούν να έχουν αναπαραστάσεις. Μια αναπαράσταση είναι μια ροή bytes σε συγκεκριμένη μορφή, όπως HTML, RDF/XML ή JPEG. Για παράδειγμα, ένα τιμολόγιο είναι ένας πληροφοριακός πόρος. Θα μπορούσε να αναπαρασταθεί ως σελίδα HTML, ως εκτυπώσιμο έγγραφο PDF ή ως έγγραφο RDF. Ένας πληροφοριακός πόρος μπορεί να έχει πολλές διαφορετικές αναπαραστάσεις, για παράδειγμα, σε διαφορετικές μορφές, ποιότητες ευκρίνειας ή φυσικές γλώσσες. ^[7]

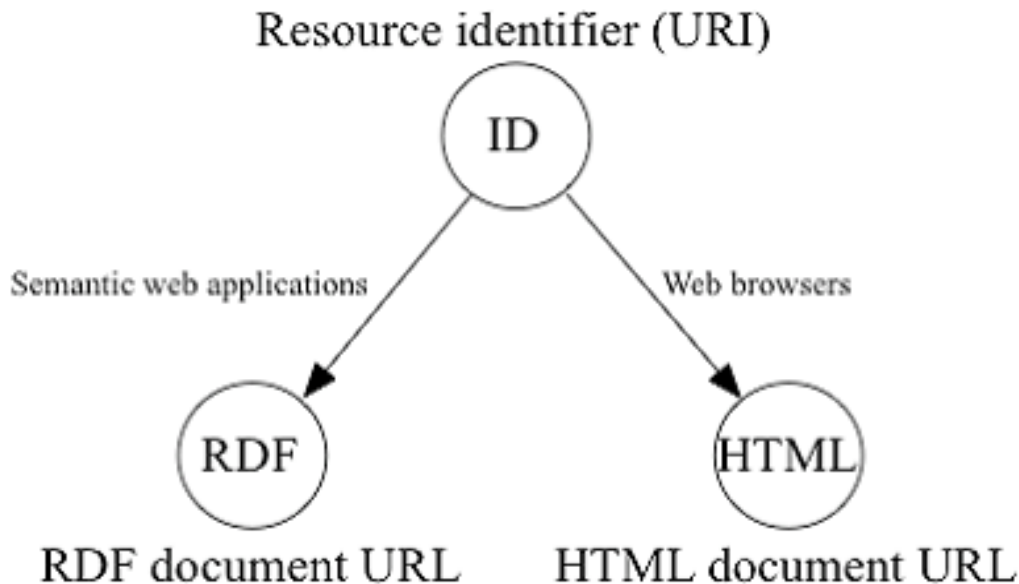
Οι πόροι ταυτοποιούνται με τη χρήση Ενιαίων Αναγνωριστικών Πόρων (URIs). Για να ανακαλύψουμε ποιον πόρο ταυτοποιεί ένα URI, θα θέλαμε να έχουμε μια υπηρεσία όπου να μπορούμε να ανατρέξουμε για την περιγραφή του URI, παρόμοια με τις σημερινές μηχανές αναζήτησης. Ωστόσο, για να συμβαδίσουμε με την αποκεντρωμένη φύση του Ιστού, θα πρέπει αντί γι' αυτό, να χρησιμοποιήσουμε τον ίδιο τον Ιστό – ένα εξαιρετικά εύρωστο και κλιμακούμενο σύστημα δημοσίευσης πληροφοριών – ως υπηρεσία στην οποία μπορούμε να ανατρέξουμε για περιγραφές πόρων. Όποτε αναφέρεται ένα URI μπορούμε να ανατρέξουμε σε αυτό ώστε να ανακτήσουμε μια περιγραφή που να περιέχει σχετικές πληροφορίες και συνδέσμους προς τα σχετιζόμενα δεδομένα. Επομένως, πολύ σημαντικό είναι τα URIs που θα δημιουργηθούν να βρίσκονται στον Ιστό.

Δεδομένου ενός **URI**, οι μηχανές θα πρέπει να λαμβάνουν **RDF δεδομένα** και οι άνθρωποι θα πρέπει να λαμβάνουν μια αναγνώσιμη αναπαράσταση, όπως είναι η HTML. Πρέπει να χρησιμοποιηθεί το πρότυπο πρωτόκολλο μεταφοράς στον Ιστό, το HTTP. Το σχήμα `http://` είναι το μοναδικό URI σχήμα το οποίο υποστηρίζεται ευρέως στα σημερινά εργαλεία και υποδομές, γι' αυτό στο πλαίσιο των Συνδεδεμένων Δεδομένων περιοριζόμαστε στη χρήση HTTP URIs για τα πάντα και αποφεύγουμε άλλα σχήματα URI. Τα HTTP URIs είναι καλά ονόματα για δύο λόγους: Παρέχουν έναν απλό τρόπο για τη δημιουργία παγκόσμια μοναδικών ονομάτων χωρίς κεντρική διαχείριση. Τα URIs λειτουργούν όχι μόνο ως ονόματα αλλά επίσης ως ένα μέσο αξιολόγησης της πληροφορίας για έναν πόρο στον Ιστό.

Μια δεύτερη προϋπόθεση για τα URIs είναι να μην υπάρχει αμφισημία μεταξύ τους, δηλαδή δεν πρέπει να υπάρχει σύγχυση μεταξύ αναγνωριστικών για έγγραφα (URLs) και αναγνωριστικών πόρων. Τα URIs προορίζονται για την αναγνώριση ενός μόνου πράγματος και ένα URI δεν μπορεί να επέχει θέση και εγγράφου ανακτήσιμου μέσω Ιστού και αντικειμένου του πραγματικού κόσμου.

Σημειώνεται ότι οι προϋποθέσεις μας φαίνονται να συγκρούονται μεταξύ τους. Εάν δεν μπορούμε να χρησιμοποιήσουμε URLs εγγράφων για την ταυτοποίηση αντικειμένων του πραγματικού κόσμου, τότε πώς θα ανακτήσουμε αντικείμενα του πραγματικού κόσμου βασιζόμενοι στα URL τους; Η πρόκληση είναι να βρεθεί μια λύση που να μας επιτρέψει να ανακαλύψουμε τα περιγραφικά έγγραφα με δεδομένο μόνο το URI του πόρου, χρησιμοποιώντας πρότυπες τεχνολογίες του Ιστού. ^[8]

Αυτό φαίνεται στο παρακάτω σχήμα:



Σχήμα 3.5 : Ένας Πόρος και τα Περιγραφικά του Εγγραφα

3.3.2 Dereferencing HTTP URI's

Η διαδικασία της «επίσκεψης» ενός URI στον Ιστό με σκοπό να ανακτήσουμε πληροφορίες σχετικά με τον αναφερόμενο πόρο λέγεται dereferencing HTTP URIs. Το προσχέδιο του W3C TAG εισήγαγε μια διάκριση στο πώς γίνεται dereferencing στα URIs που ταυτοποιούν πληροφοριακούς πόρους και μη-πληροφοριακούς πόρους:

- **Πληροφοριακοί Πόροι:** Όταν επισκεφθούμε ένα URI που ταυτοποιεί έναν πληροφοριακό πόρο, ο εξυπηρετητής του ιδιοκτήτη του URI συνήθως παράγει μια νέα αναπαράσταση, μια νέα στιγμιαία αναπαράσταση της τρέχουσας κατάστασης της πληροφοριακής πηγής και τη στέλνει πίσω στον πελάτη χρησιμοποιώντας τον HTTP κωδικό απόκρισης 200 OK.
- Δεν μπορούμε να επισκεφθούμε **τους μη-πληροφοριακούς πόρους** απ' ευθείας. Επομένως, η αρχιτεκτονική του Ιστού χρησιμοποιεί ένα κόλπο για να επιτρέψει την επίσκεψη στα URIs που

ταυτοποιούν μη-πληροφοριακούς πόρους: αντί να αποστέλλεται μια αναπαράσταση του πόρου, ο εξυπηρετητής στέλνει στον πελάτη το URI ενός πληροφοριακού πόρου που περιγράφει τον μη-πληροφοριακό πόρο, χρησιμοποιώντας τον HTTP κωδικό απόκρισης 303 See Other. Αυτό ονομάζεται 303 ανακατεύθυνση. Σε ένα δεύτερο βήμα, ο πελάτης επισκέπτεται αυτό το νέο URI και παίρνει μια αναπαράσταση που περιγράφει τον αρχικό μη-πληροφοριακό πόρο. ^[7]

- Η διαδικασία αυτή καθίσταται δυνατή χάρη σε έναν HTTP μηχανισμό που ονομάζεται **διαπραγμάτευση περιεχομένου** (content negotiation).

Σημείωση 1: Υπάρχουν δύο προσεγγίσεις για το πώς μπορούν οι εκδότες δεδομένων να παρέχουν στους πελάτες URIs πληροφοριακών πόρων που να περιγράφουν μη-πληροφοριακούς πόρους: τα Hash (δίεση) URIs και τα 303 URIs .

Σημείωση 2: Κάποιος θα μπορούσε να σκεφτεί ότι πρέπει με κάποιο τρόπο να οριοθετήσουμε τα παραδοσιακά έγγραφα του Ιστού από τους άλλους, μη-περιγραφικούς πόρους. Σύμφωνα με τις οδηγίες του W3C, μπορούμε να έχουμε έγγραφο του Ιστού εάν όλα τα βασικά χαρακτηριστικά του μπορούν να διαβιβαστούν μέσα σε ένα μήνυμα. Επειδή αυτός ο ορισμός δεν είναι και πολύ ακριβής, συστήνεται να σφάλουμε επί το ασφαλέστερον: όταν το αντικείμενο του ενδιαφέροντός μας δεν είναι εμφανώς και προφανώς έγγραφο, τότε είναι καλύτερο να χρησιμοποιούμε δύο διακριτά URIs, ένα για τον πόρο και άλλο ένα για το έγγραφο που τον περιγράφει. ^[8]

3.3.3 Διαπραγμάτευση Περιεχομένου (content negotiation).

Οι HTML φυλλομετρητές συνήθως εμφανίζουν τις RDF αναπαραστάσεις ως «γυμνό» κώδικα RDF ή απλά τις «κατεβάζουν» ως αρχεία RDF χωρίς να τις εμφανίζουν. Αυτό δεν είναι πολύ βοηθητικό για το μέσο χρήστη. Επομένως, η παρουσίαση μιας σωστής HTML αναπαράστασης σε συνδυασμό με την RDF αναπαράσταση ενός πόρου βοηθάει τον άνθρωπο να καταλάβει σε τί αναφέρεται ένα URI.

Επομένως, διαπραγμάτευση περιεχομένου είναι η διαδικασία επιλογής της καλύτερης αναπαράστασης για μια δεδομένη απόκριση όταν υπάρχουν διαθέσιμες πολλαπλές αναπαραστάσεις. Οι πελάτες HTTP στέλνουν κεφαλίδες HTTP με κάθε αίτηση για να δηλώσουν

τί είδος αναπαράστασης προτιμούν. Εάν οι κεφαλίδες δηλώνουν ότι ο πελάτης προτιμά HTML, τότε ο εξυπηρετητής μπορεί να παράγει μια HTML αναπαράσταση. Εάν ο πελάτης προτιμά RDF, τότε ο εξυπηρετητής μπορεί να παράγει RDF.

Παράδειγμα 1 :

```
curl -I -H "Accept: application/rdf+xml" http://dbpedia.org/resource/Galway
```

```
HTTP/1.1 303 See Other
```

```
Content-Type: application/rdf+xml
```

```
Location: http://dbpedia.org/data/Galway.rdf
```

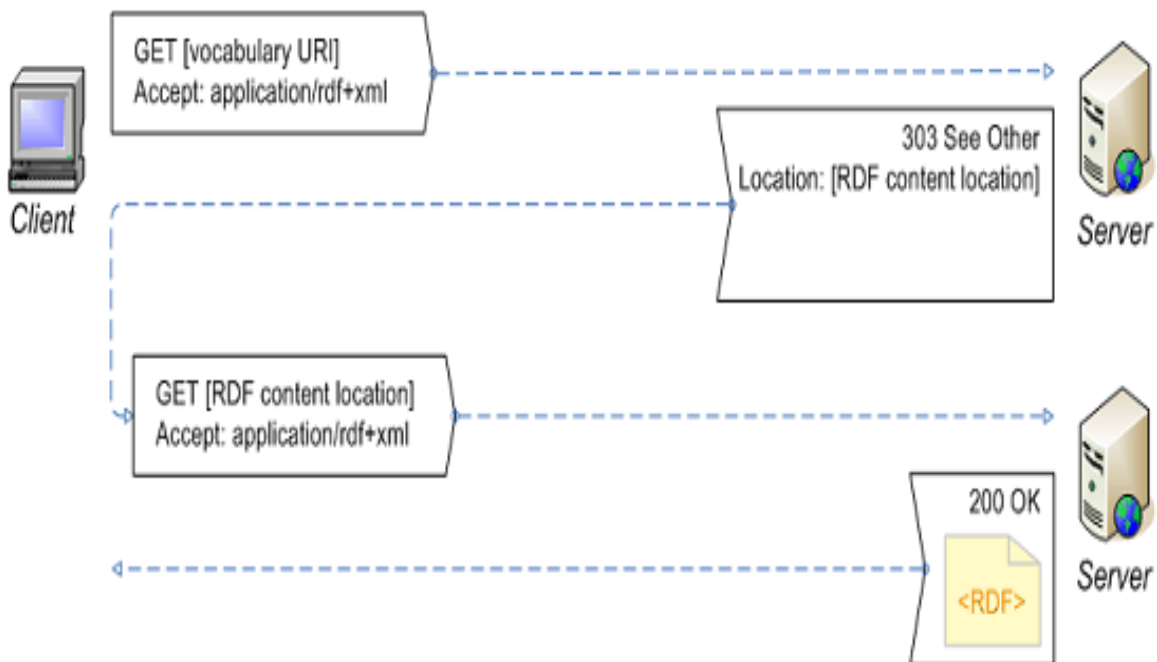
Η διαπραγμάτευση περιεχομένου για μη-πληροφοριακούς πόρους υλοποιείται κυρίως με τον ακόλουθο τρόπο. Όταν επισκεφτούμε ένα URI που ταυτοποιεί ένα μη-πληροφοριακό πόρο, ο εξυπηρετητής στέλνει μια ανακατεύθυνση 303 προς έναν πληροφοριακό πόρο κατάλληλο για τον πελάτη. Επομένως, μια πηγή δεδομένων συχνά εξυπηρετεί τρία URIs που σχετίζονται με μη-πληροφοριακούς πόρους, για παράδειγμα:

- <http://www4.wiwiss.fu-berlin.de/factbook/resource/Russia> (URI που ταυτοποιεί το μη-πληροφοριακό πόρο Ρωσία)
- <http://www4.wiwiss.fu-berlin.de/factbook/data/Russia> (πληροφοριακός πόρος με μια RDF/XML αναπαράσταση που περιγράφει τη Ρωσία)
- <http://www4.wiwiss.fu-berlin.de/factbook/page/Russia> (πληροφοριακός πόρος με μια HTML αναπαράσταση που περιγράφει τη Ρωσία)

Η παρακάτω εικόνα δείχνει πώς η επίσκεψη σε ένα HTTP URI το οποίο ταυτοποιεί ένα μη-πληροφοριακό πόρο χρησιμοποιεί τη διαπραγμάτευση περιεχομένου:

1. Ο πελάτης πραγματοποιεί ένα HTTP GET αίτημα προς ένα URI το οποίο ταυτοποιεί ένα μη-πληροφοριακό πόρο. Στην περίπτωσή μας, ένα URI λεξιλογίου. Εάν ο πελάτης είναι φυλλομετρητής Συνδεδεμένων Δεδομένων και προτιμά μια 26RDF/XML αναπαράσταση του πόρου, στέλνει μια κεφαλίδα *Accept: application/rdf+xml* μαζί με την αίτηση. Οι HTML φυλλομετρητές θα έστελναν αντί γι' αυτό μια κεφαλίδα *Accept: text/html*.

2. Ο εξυπηρετητής αναγνωρίζει το URI που ταυτοποιεί ένα μη-πληροφοριακό πόρο. Καθώς ο εξυπηρετητής δεν μπορεί να επιστρέψει μια αναπαράσταση αυτού του πόρου, απαντά χρησιμοποιώντας τον HTTP κωδικό απόκρισης 303 See Other και στέλνει στον πελάτη το URI ενός πληροφοριακού πόρου που περιγράφει το μη-πληροφοριακό πόρο. Στην περίπτωση του RDF: *RDF content location*.
3. Ο πελάτης τώρα ζητά από τον εξυπηρετητή να λάβει (GET) μια αναπαράσταση του πληροφοριακού πόρου, ζητώντας ξανά *application/rdf+xml*.
4. Ο εξυπηρετητής στέλνει στον πελάτη ένα έγγραφο RDF/XML το οποίο περιέχει μια περιγραφή του *vocabulary URI* του αρχικού πόρου.



Σχήμα 3.6 : Διαπραγμάτευση Περιεχομένου

3.3.4 Ταυτόσημα URI's (URI aliases)

Σε ένα ανοιχτό περιβάλλον όπως είναι ο Ιστός συχνά συμβαίνει διαφορετικοί πάροχοι να μιλούν για τον ίδιο μη-πληροφοριακό πόρο, για παράδειγμα μια γεωγραφική τοποθεσία ή ένα διάσημο πρόσωπο. Καθώς μπορεί να μη γνωρίζουν ο ένας για τον άλλο, συνιστούν διαφορετικά URIs για να ταυτοποιήσουν την ίδια οντότητα. Για παράδειγμα, η DBpedia, μια πηγή δεδομένων που παρέχει πληροφορίες οι οποίες έχουν εξαχθεί από τη Wikipedia, χρησιμοποιεί το URI <http://dbpedia.org/resource/Berlin> για να ταυτοποιήσει το Βερολίνο. Το Geonames, μια πηγή δεδομένων που παρέχει πληροφορίες για εκατομμύρια γεωγραφικές τοποθεσίες, χρησιμοποιεί το URI <http://sws.geonames.org/2950159/> για να ταυτοποιήσει το Βερολίνο. Καθώς και τα δύο URIs αναφέρονται στον ίδιο μη-πληροφοριακό πόρο, καλούνται **ταυτόσημα URI**.

Τα ταυτόσημα URIs συνηθίζονται στον Ιστό των Δεδομένων καθώς δεν μπορεί να περιμένουμε ότι όλοι οι πάροχοι πληροφοριών θα συμφωνήσουν πάνω στο ίδιο URI για να ταυτοποιήσουν μια οντότητα. Τα ταυτόσημα URIs παρέχουν μια σημαντική κοινωνική λειτουργία στον Ιστό Δεδομένων καθώς αναφέρονται σε διαφορετικές περιγραφές του ίδιου μη-πληροφοριακού πόρου και έτσι επιτρέπουν την έκφραση διαφορετικών όψεων και απόψεων. Για να είμαστε ακόμα σε θέση να εντοπίσουμε ότι οι διαφορετικοί πάροχοι πληροφοριών μιλούν για τον ίδιο μη-πληροφοριακό πόρο, είναι κοινή πρακτική για τους παρόχους πληροφοριών να θέτουν συνδέσμους `owl:sameAs` προς ταυτόσημα URI τα οποία γνωρίζουν.

3.3.5 Σχετιζόμενες Περιγραφές

Ένας χρήσιμος όρος στα πλαίσια των Συνδεδεμένων Δεδομένων (αλλά όχι μέρος της καθορισμένης ορολογίας της Αρχιτεκτονικής Ιστού) είναι η σχετιζόμενη περιγραφή και αναφέρεται στην περιγραφή ενός μη-πληροφοριακού πόρου την οποία αποκτά ο πελάτης επισκεπτόμενος ένα συγκεκριμένο URI που ταυτοποιεί αυτό το μη-πληροφοριακό πόρο. Για παράδειγμα, επισκεπτόμενοι το URI <http://dbpedia.org/resource/Berlin> και ζητώντας *application/rdf+xml*, μας δίνεται, μετά την ανακατεύθυνση, μια σχετιζόμενη περιγραφή που είναι ίδια με την RDF περιγραφή του <http://dbpedia.org/resource/Berlin> μέσα στον πληροφοριακό πόρο <http://dbpedia.org/data/Berlin>. Η χρήση αυτού του νέου όρου έχει νόημα στα πλαίσια των Συνδεδεμένων Δεδομένων καθώς είναι συνηθισμένη η χρήση πολλαπλών ταυτόσημων URI για

την αναφορά στον ίδιο μη-πληροφοριακό πόρο και επίσης διότι διαφορετικά ταυτόσημα URI αναφέρονται σε διαφορετικές περιγραφές του πόρου.

3.4 Linking Open Data Project

Το πιο εμφανές παράδειγμα υιοθέτησης και εφαρμογής των κανόνων των Συνδεδεμένων Δεδομένων είναι το project «Σύνδεση Ανοιχτών Δεδομένων» (Linking Open Data - LOD), μια κοινοτική προσπάθεια που θεμελιώθηκε τον Ιανουάριο του 2007 από τους Chris Bizer και Richard Cyganiak και υποστηρίχθηκε από το W3C Semantic Web Education και το Outreach Group. Το εγχείρημα αυτό ενθαρρύνει τους ανθρώπους να αναφέρονται σε συγκεκριμένα αντικείμενα (όπως είναι μια ταινία, ένα πρόσωπο ή ένα εστιατόριο) με ένα πρότυπο Κοινό Αναγνωριστικό Πόρου (URI), έτσι ώστε ο καθένας να ξέρει πότε δύο άνθρωποι μιλάνε για το ίδιο αντικείμενο. Ο αρχικός και τεκταινόμενος σκοπός του project είναι να συνυφάνει τον Ιστό των Δεδομένων αναγνωρίζοντας τα υπάρχοντα σύνολα δεδομένων, τα οποία διατίθενται ανοιχτά, μετατρέποντάς τα σε RDF σύμφωνα με τους κανόνες των Συνδεδεμένων Δεδομένων και δημοσιεύοντάς τα στον Ιστό.

Οι συμμετέχοντες στα πρώιμα στάδια του project ήταν κυρίως ερευνητές και προγραμματιστές σε πανεπιστημιακά ερευνητικά εργαστήρια και μικρές εταιρίες.

Τα μεγαλύτερα διασυνδεδεμένα σύνολα δεδομένων του Ιστού ήταν αρχικά η DBpedia¹⁰ (91 εκατομμύρια τριάδες), το Geonames¹¹ (60 εκατομμύρια τριάδες), το Musicbrainz¹² (50 εκατομμύρια τριάδες), ο DBtune music serve¹³ (4 εκατομμύρια τριάδες), η βιβλιογραφία DBLP¹⁴ (15 εκατομμύρια τριάδες), οι αξιολογήσεις Revyu¹⁵ (15 χιλιάδες τριάδες), το σύνολο δεδομένων απογραφής των ΗΠΑ (700 εκατομμύρια τριάδες) και το RDF Book Mashup¹⁶ (αρκετά δισεκατομμύρια τριάδες).

¹⁰ <http://dbpedia.org>

¹¹ <http://www.geonames.org/>

¹² <http://dbtune.org/musicbrainz/>

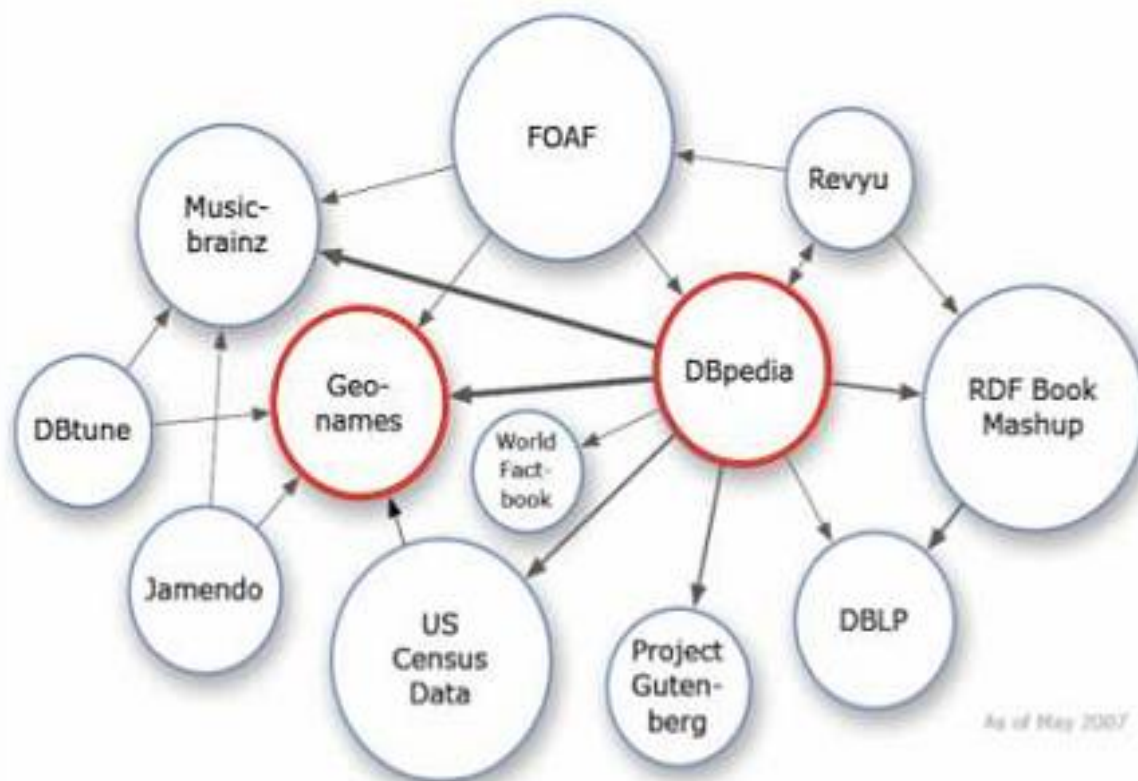
¹³ <http://dbtune.org/>

¹⁴ <http://www4.wiwiss.fu-berlin.de/dblp/>

¹⁵ <http://revyu.com/>

¹⁶ <http://www4.wiwiss.fu-berlin.de/bizer/bookmashup/>

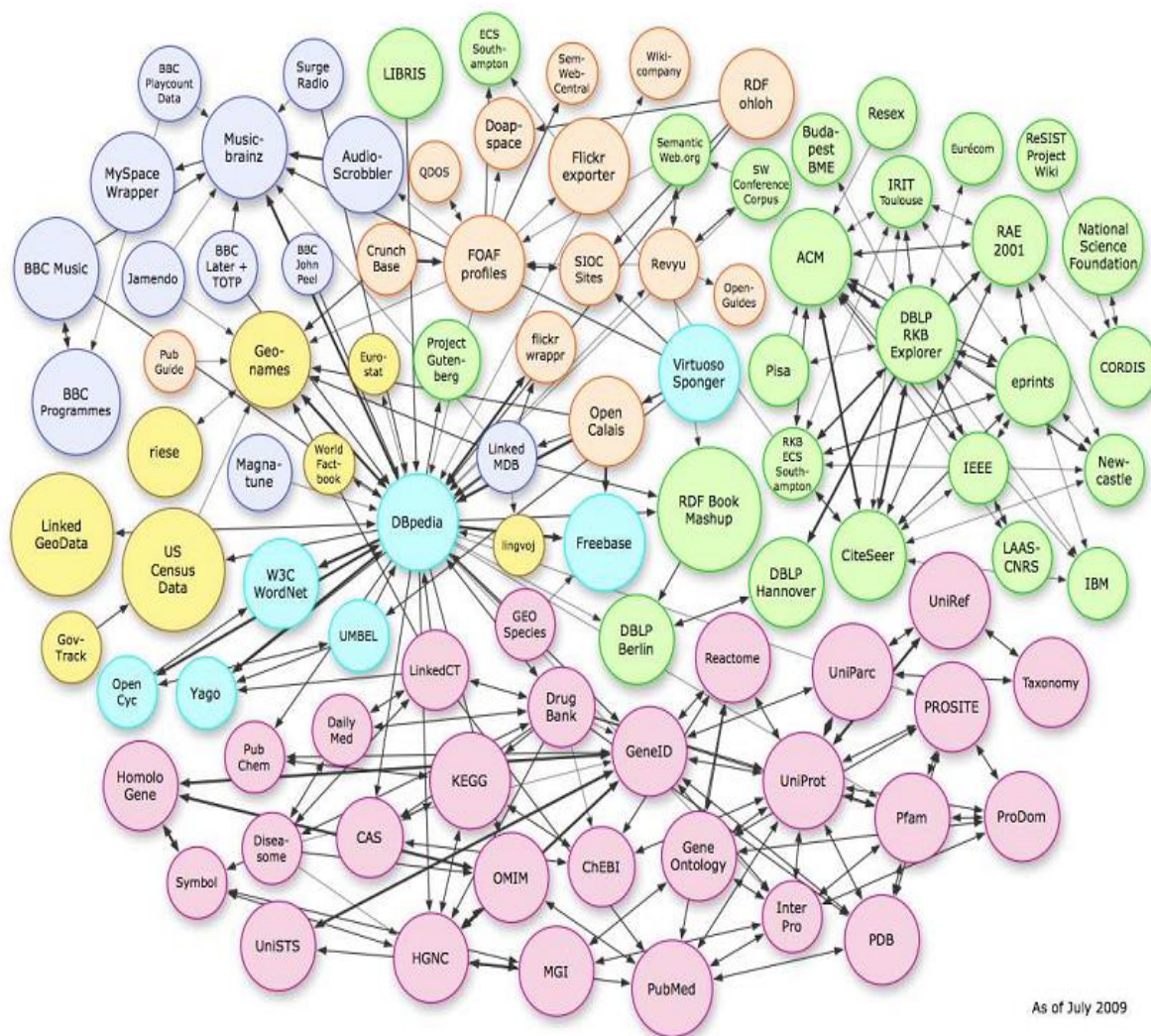
Αυτά τα σύνολα δεδομένων διασυνδέονταν με περίπου 150.000 RDF συνδέσμους, στη μορφή τριάδων που συνδέουν ένα υποκείμενο URI από ένα σύνολο δεδομένων με ένα αντικείμενο URI από ένα άλλο σύνολο δεδομένων. Χρησιμοποιώντας αυτούς τους συνδέσμους μπορούσε κάποιος να πλοηγηθεί από έναν επιστήμονα υπολογιστών στη DBpedia στις δημοσιεύσεις του στη βάση δεδομένων DBLP, από ένα βιβλίο της DBpedia σε αξιολογήσεις και προσφορές για το βιβλίο αυτό που παρέχονται από το RDF Book Mashup ή από ένα συγκρότημα στη DBpedia στη λίστα τραγουδιών τους που παρέχεται από το Musicbrainz ή το DBtune.



Σχήμα 3.7 : Σύννεφο Linking Open Data Project , Μαΐος 2007

Έκτοτε το project αναπτύχθηκε σημαντικά ώστε να περιλάβει σημαντικές συμμετοχές από μεγάλους οργανισμούς όπως το BBC, το Thomson Reuters και τη Βιβλιοθήκη του Κογκρέσου. Αυτή η ανάπτυξη καταστάθηκε εφικτή από την ανοιχτή φύση του project, καθώς μπορεί να συμμετάσχει ο καθένας, δημοσιεύοντας απλά ένα σύνολο δεδομένων σύμφωνα με τους κανόνες των Συνδεδεμένων Δεδομένων και διασυνδέοντάς το με τα υπάρχοντα σύνολα δεδομένων.

Ενδεικτικό του εύρους και της κλίμακας του Ιστού Δεδομένων που προήλθε από το Linking Open Data project είναι το Σχήμα 3.8. Κάθε κόμβος στο διάγραμμα σύννεφο αναπαριστά ένα ξεχωριστό σύνολο δεδομένων που δημοσιεύεται ως Συνδεδεμένα Δεδομένα.



Σχήμα 3.8 : Διάγραμμα – Σύννεφο των Linking Open Data που δίνει μια γενική εικόνα των δημοσιευμένων συνόλων δεδομένων και τις σχέσεις τους

Τα βέλη στο Σχήμα 3.8 υποδεικνύουν ότι υπάρχουν σύνδεσμοι μεταξύ αντικειμένων στα δύο συνδεόμενα σύνολα δεδομένων. Τα πιο «παχιά» βέλη αντιστοιχούν σε μεγαλύτερο αριθμό συνδέσμων μεταξύ δύο συνόλων δεδομένων, ενώ τα αμφίδρομα βέλη υποδεικνύουν ότι σε κάθε σύνολο δεδομένων υπάρχουν εξερχόμενοι σύνδεσμοι προς το άλλο.

Το περιεχόμενο του σύννεφου έχει ποικίλη φύση, συνδυάζοντας δεδομένα από γεωγραφικές τοποθεσίες, εταιρίες, βιβλία, επιστημονικές δημοσιεύσεις, ταινίες, μουσική, τηλεοπτικά και ραδιοφωνικά προγράμματα, γονίδια, πρωτεΐνες, φάρμακα και κλινικές δοκιμές, διαδικτυακές κοινότητες, στατιστικά δεδομένα, δημογραφικά αποτελέσματα και αξιολογήσεις.

Ο υπολογισμός του ακριβούς μεγέθους του Ιστού των Δεδομένων αποτελεί πρόκληση εξαιτίας του γεγονότος ότι πολλά από τα δεδομένα έχουν παραχθεί από wrappers γύρω από υπάρχουσες σχεσιακές βάσεις δεδομένων ή APIs και επομένως απαιτείται πρώτα να «περπατηθούν» πριν να μπορέσουν να καταμετρηθούν ή να αναλυθούν. Εναλλακτικά, το μέγεθος του Ιστού Δεδομένων μπορεί να εκτιμηθεί με βάση τα στατιστικά των συνόλων δεδομένων που έχουν συλλεχθεί από την LOD κοινότητα στο ESW wiki. Σύμφωνα με αυτά τα στατιστικά¹⁷, ο Ιστός των Δεδομένων αποτελείται σήμερα από 4,7 δισεκατομμύρια τριάδες, οι οποίες διασυνδέονται με περίπου 142 εκατομμύρια RDF συνδέσμους (Μάιος 2009).

Όπως δείχνει το Σχήμα 3.8, μερικά σύνολα δεδομένων χρησιμεύουν ως συνδετικοί κόμβοι στον Ιστό των Δεδομένων. Για παράδειγμα, το σύνολο δεδομένων της DBpedia αποτελείται από RDF τριάδες που έχουν αποσπασθεί από τα «infoboxes» που υπάρχουν συνήθως στη δεξιά πλευρά των άρθρων της Wikipedia, ενώ το Geonames παρέχει RDF περιγραφές εκατομμυρίων γεωγραφικών τοποθεσιών παγκοσμίως. Καθώς αυτά τα δύο σύνολα δεδομένων παρέχουν URIs και RDF περιγραφές για πολλές συνηθισμένες οντότητες ή έννοιες, συχνά αναφέρονται και σε άλλα πιο εξειδικευμένα σύνολα δεδομένων και έτσι έχουν εξελιχθεί σε κόμβους στους οποίους συνδέονται όλο και περισσότερα σύνολα δεδομένων.

Τα σύνολα δεδομένων RDF που δημιουργήθηκαν από το project μπορούν να χρησιμοποιηθούν ως δοκιμαστικά από φυλλομετρητές του Σημασιολογικού Ιστού, αποθήκες RDF και μηχανές συλλογιστικής (reasoning), καθώς και για σύνδεση δεδομένων, καθαρισμό δεδομένων και εργαλεία εξόρυξης δεδομένων. Τα άτομα ενθαρρύνονται να θέτουν RDF συνδέσμους μέσα στα σύνολα δεδομένων τους, καθώς κάθε νέος σύνδεσμος βοηθά στη συνύφανση του Σημασιολογικού Ιστού ως όλου.^[10]

Έχουν γίνει επίσης αρκετές προσπάθειες να προτυποποιηθεί ένα σύνολο οντολογιών, το οποίο να περιγράφει ποια πεδία θα πρέπει να χρησιμοποιούνται για να περιγράψουν πράγματα όπως ένα εστιατόριο ή μια ταινία σε όλες τις περιπτώσεις. Μέχρι σήμερα, ωστόσο, ο αριθμός των ομάδων που έχουν συμφωνήσει να χρησιμοποιούν το ίδιο URI για να αναφερθούν σε αντικείμενα και τις ίδιες οντολογίες για να τα περιγράψουν αποτελεί ένα εξαιρετικά μικρό ποσοστό όλων των

¹⁷ <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets/LinkStatistics>
και <http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/DataSets/Statistics>

διαθέσιμων online βάσεων δεδομένων. Σε πολλές περιπτώσεις, ακόμα και αυτοί που προσπαθούν να συμμετάσχουν στις προσπάθειες των Συνδεδεμένων Ανοιχτών

3.5 Συνδεδεμένα Κυβερνητικά Ανοιχτά δεδομένα

3.5.1 Δημοσίευση Κυβερνητικών Δεδομένων στον Ιστό ως Συνδεδεμένα

Δεδομένα

Δεδομένων δε χρησιμοποιούν στην παρούσα φάση τα ίδια URIs για αντικείμενα τα οποία αναφέρονται ξεκάθαρα στο ίδιο πράγμα. Αυτό σημαίνει ότι εάν επιθυμούμε να συνδέσουμε σύνολα δεδομένων, θα πρέπει να επινοήσουμε τρόπους για να αποφασίζεται με αυτόματο τρόπο εάν δύο αντικείμενα είναι ίδια.^[9]

Με μια πιο προσεκτική ματιά στα ευρέως χρησιμοποιούμενα λεξιλόγια στα συνδεδεμένα σύνολα δεδομένων κάποιος μπορεί να ομαδοποιήσει τους «τύπους σημασιολογικών συνδέσμων» όπως για παράδειγμα:

- Σύνδεσμοι σχετιζόμενοι με πρόσωπα, για παράδειγμα foaf:knows από FOAF
- Χωρικοί τύποι συνδέσμων όπως foaf:based_near ή geo:lat από το βασικό geo (WGS84 lat/long) λεξιλόγιο
- Χρονικοί τύποι δεδομένων, όπως η ιδιότητα dc:created του Dublin Core και η ιδιότητα event:time του Event Ontology
- Τύποι συνδέσμων όπως dc:isPartOf για την αναπαράσταση δομικών σημασιολογιών
- Άλλοι, όπως ο scovo:dimension του Statistical Core Vocabulary.^[5]

Τα Συνδεδεμένα Δεδομένα αποτελούν την πιο κατάλληλη λύση για τη δημοσίευση ανοιχτών κυβερνητικών δεδομένων στον Ιστό, αφού οι κανόνες τους συνάδουν με τους κανόνες των ανοιχτών κυβερνητικών δεδομένων, όπως παρουσιάστηκαν παραπάνω. Η χρήση των URIs και του RDF, που αποτελούν παγκόσμια προσβάσιμες μορφές, συνεπάγεται την τυποποίηση της ταυτοποίησης και αναπαράστασης των κυβερνητικών πόρων. Με τον τρόπο αυτό, θα μειωθούν τα εμπόδια στη σύνδεση ανοιχτών κυβερνητικών δεδομένων σε παγκόσμια κλίμακα.^[19]

Ο Tim Berners-Lee, σε αναφορά του σχετικά με τη δημοσίευση Κυβερνητικών Δεδομένων στον Ιστό, επισημαίνει ότι «κατ' αρχήν, τα πρωτογενή δεδομένα πρέπει να γίνουν διαθέσιμα στον Ιστό το συντομότερο δυνατό. Κατά προτίμηση, πρέπει να δημοσιευτούν ως Συνδεδεμένα Δεδομένα. Το επόμενο βήμα είναι να συνδεθούν με άλλες πηγές. Ως χαμηλότερη προτεραιότητα, πρέπει να δημιουργηθούν ωραίες διεπαφές χρήστη – εάν οι ενδιαφερόμενες κοινότητες έξω από την κυβέρνηση δεν το έχουν κάνει ήδη. Η τεχνολογία των Συνδεδεμένων Δεδομένων, αντίθετα με κάθε άλλη τεχνολογία, επιτρέπει σε κάθε επικοινωνία δεδομένων να αποτελείται από πολλά αναμειγμένα λεξιλόγια. Κάθε λεξιλόγιο προέρχεται από μια κοινότητα είτε είναι διεθνής, εθνική, κρατική ή τοπική, ή εξειδικευμένη σε κάποιο τομέα της βιομηχανίας.

Τα κυβερνητικά δεδομένα τοποθετούνται δικτυακά για τρεις κυρίως λόγους:

1. Αυξάνουν την επίγνωση του πολίτη για τις κυβερνητικές λειτουργίες ώστε να επιτευχθεί μεγαλύτερη αξιοπιστία
2. Συμβάλλουν με πολύτιμες πληροφορίες που αφορούν τον κόσμο, και
3. Δίνουν τη δυνατότητα στην κυβέρνηση, τη χώρα και τον κόσμο να λειτουργήσουν πιο αποδοτικά.

Καθένας από τους σκοπούς αυτούς εξυπηρετείται με βέλτιστο τρόπο με τη χρήση τεχνικών των Συνδεδεμένων Δεδομένων. Γενικά τα Συνδεδεμένα Δεδομένα είναι:

Ανοιχτά: Τα Συνδεδεμένα Δεδομένα είναι προσβάσιμα μέσω απεριόριστης ποικιλίας εφαρμογών διότι είναι εκφρασμένα σε ανοιχτά, μη-ιδιωτικά σχήματα.

Modular: Τα Συνδεδεμένα Δεδομένα μπορούν να **συνδυαστούν** (πολτοποιηθούν) με κάθε άλλο κομμάτι Συνδεδεμένων Δεδομένων. Για παράδειγμα, τα κυβερνητικά δεδομένα για τα έξοδα της φροντίδας υγείας για μια δεδομένη γεωγραφική περιοχή μπορούν να συνδυαστούν με άλλα δεδομένα για τα χαρακτηριστικά του πληθυσμού αυτής της περιοχής ώστε να αξιολογηθεί η αποτελεσματικότητα των κυβερνητικών προγραμμάτων. Δεν απαιτείται προηγούμενος σχεδιασμός για το συγκερασμό αυτών των πηγών δεδομένων εφόσον χρησιμοποιούν και οι δύο τα πρότυπα των Συνδεδεμένων Δεδομένων.

Κλιμακούμενα: Είναι εύκολο να προσθέσουμε επιπλέον Συνδεδεμένα Δεδομένα σε αυτά που ήδη υπάρχουν, ακόμα και όταν οι όροι και οι ορισμοί που χρησιμοποιούνται αλλάζουν με το χρόνο.

Το ουσιαστικό μήνυμα είναι ότι σε οποιαδήποτε μορφή κι αν επιθυμούν τα δεδομένα τους οι άνθρωποι και, σε οποιαδήποτε μορφή κι αν τα διαθέτουν, το μοντέλο RDF πρέπει να χρησιμοποιηθεί ως συνδετικός κρίκος. Όταν τα δεδομένα δημοσιεύονται σε μορφή RDF, καθίστανται διαθέσιμα για επεξεργασία όχι απλά διαθέσιμα για δημόσια πρόσβαση. Με το RDF, ένας φυλλομετρητής Ιστού θα μπορεί να επαναχρησιμοποιήσει τα δεδομένα, μη απαιτώντας επιπλέον εργασία εκ μέρους των χρηστών. Το RDF συνδέει καλύτερα από κάθε άλλο μοντέλο διότι:

- Χρησιμοποιεί URIs, επιτρέποντας έτσι τη σύνδεση αντικειμένων και εννοιών
- Επιτρέπει σε ξεχωριστά συστήματα που είναι σχεδιασμένα ανεξάρτητα να συνενωθούν στη συνέχεια
- Επιτρέπει να προστεθεί διαλειτουργικότητα όπου είναι επωφελής
- Επιτρέπει σε όλα τα δεδομένα να εκφραστούν ως μίξη λεξιλογίων.»^[12]

Όταν σε ομιλία του στον οργανισμό TED (Φεβρουάριος 2009) ο Tim Berners-Lee εισηγήθηκε την ιδέα των Συνδεδεμένων Δεδομένων, δήλωσε ότι επιθυμεί από τις κυβερνήσεις να τεθούν επικεφαλής της προσπάθειας για το χτίσιμο του Ιστού των Δεδομένων. Αναγνώρισε την κυβέρνηση των ΗΠΑ ως υποψήφια για την πρώτη εφαρμογή αυτής της ιδέα και είναι γεγονός πως το 2009 ήταν το έτος για την τοποθέτηση κυβερνητικών δεδομένων δικτυακά, αφού μάλιστα οι κυβερνήσεις των ΗΠΑ και του Ηνωμένου Βασιλείου δεσμεύτηκαν δημόσια προς τα ανοιχτά δεδομένα.

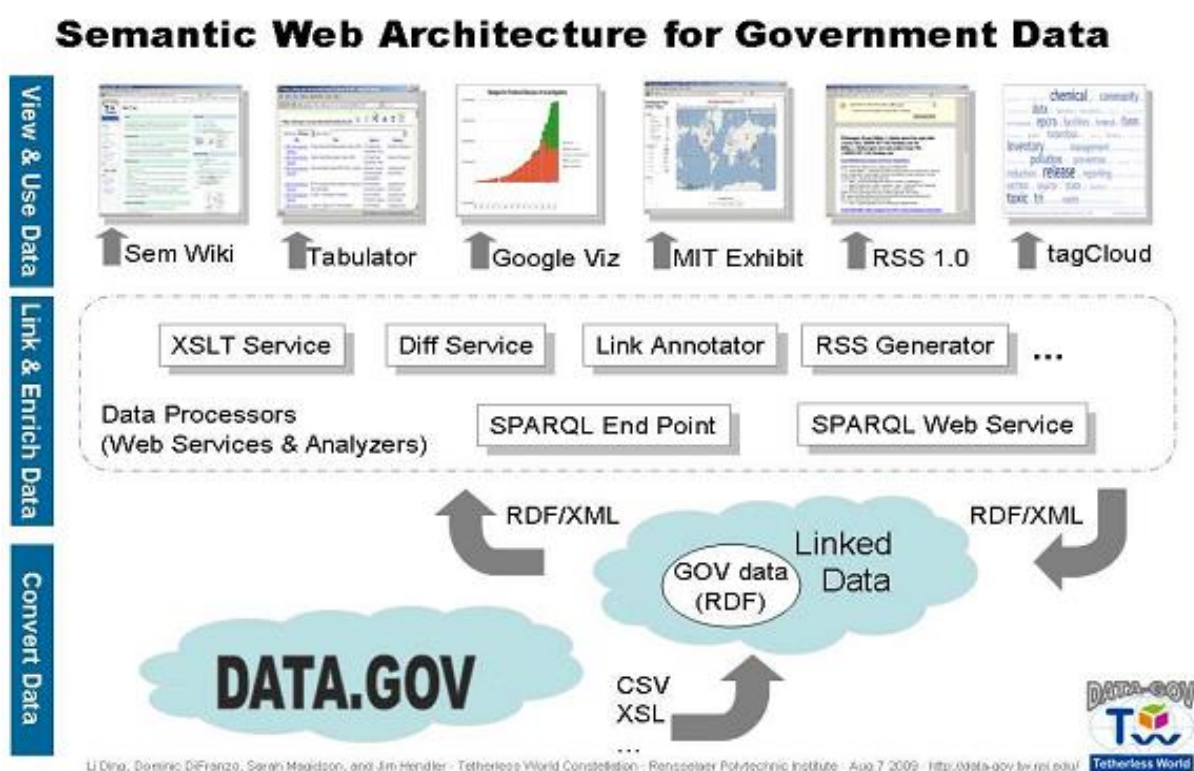
Μάλιστα, υπάρχει μια σημαντική σύνδεση μεταξύ του Linking Open Data project (LOD) και της Οδηγίας του Προέδρου των ΗΠΑ, Obama, για τη Διαφάνεια και την Ανοιχτή Διακυβέρνηση (Transparency and Open Government Directive – TODG¹⁸). Κατ' αρχήν, οι αρχές σχεδιασμού του Διαδικτύου μοιράζονται ένα κοινό σύνολο αξιών με την TODG. Δεύτερον, το LOD είναι ο μόνος τρόπος για να επιτευχθούν οι τρεις στόχοι της TODG – διαφάνεια, συμμετοχή και συνεργατικότητα – σε κλίμακα Διαδικτύου. Διότι απαιτείται το Διαδίκτυο για να παράσχει μια αρκετά ευρεία πρόσβαση στη σημερινή κοινωνία ώστε να υλοποιηθεί η TODG. Το LOD είναι μια ανερχόμενη περιοχή πρακτικής για το Σημασιολογικό Ιστό, όσον αφορά τη μεθοδολογία και τα σχήματα σχεδιασμού ή ακόμα καλύτερα, συνιστά μια σχηματική γλώσσα.^[17]

¹⁸ <http://www.whitehouse.gov/open/documents/open-government-directive>

Μερικά κυβερνητικά δεδομένα έχουν ήδη φτάσει στο σύννεφο των Συνδεδεμένων Δεδομένων – εξωτερικοί παράγοντες έχουν τροφοδοτήσει το σύννεφο με δεδομένα από το Data.gov και τη FEA (Federal Enterprise Architecture) σε RDF.

Όσον αφορά τα δεδομένα από το Data.gov, έχει αναπτυχθεί ένα project, το Data-gov Wiki¹⁹, το οποίο ερευνά το ρόλο των τεχνολογιών του Σημασιολογικού Ιστού, και πιο συγκεκριμένα των Συνδεδεμένων Δεδομένων στην παραγωγή, επεξεργασία και χρήση των κυβερνητικών δεδομένων τα οποία προέρχονται από το data.gov. Το εγχείρημα αυτό περιλαμβάνει τη μετάφραση των κυβερνητικών συνόλων δεδομένων σε RDF, τη σύνδεσή τους με το σύννεφο των Συνδεδεμένων.

Δεδομένων και την ανάπτυξη εφαρμογών και demos πάνω στα σημασιολογικά συνδεδεμένα κυβερνητικά δεδομένα. Στο διάγραμμα που ακολουθεί παρουσιάζεται η αρχιτεκτονική του Σημασιολογικού Ιστού για κυβερνητικά δεδομένα, όπως εφαρμόζεται στην περίπτωση του Data.gov.

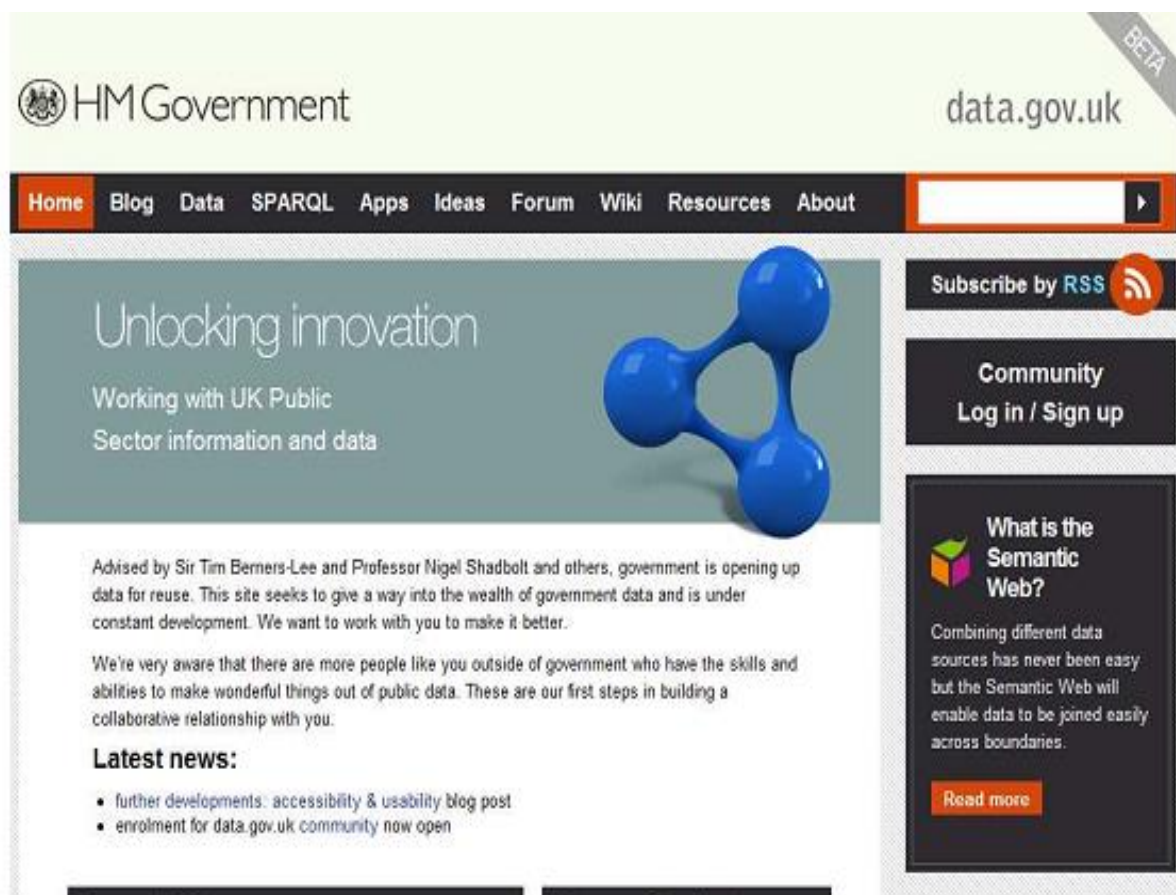


Σχήμα 3.9 : Αρχιτεκτονική Σημασιολογικού Ιστού για Κυβερνητικά Δεδομένα (Πηγή: http://data.gov.tw.rpi.edu/wiki/About_the_Data-gov_Wiki)

¹⁹ http://data.gov.tw.rpi.edu/wiki/About_the_Data-gov_Wiki

Επίσης, το Ηνωμένο Βασίλειο, θέλοντας να υλοποιήσει την Ευρωπαϊκή Οδηγία για την επαναχρησιμοποίηση των πληροφοριών του Δημοσίου Τομέα σκοπεύει να εγκαθιδρύσει ένα σημείο πρόσβασης για όλα τα δημόσια σύνολα δεδομένων, το data.gov.uk, το οποίο είναι αντίστοιχο με το data.gov των ΗΠΑ. Το site θα υλοποιείται με Drupal και θα φιλοξενεί μια αποθήκη δεδομένων χρησιμοποιώντας την πλατφόρμα Talis²⁰ [24]. Το έργο ανέλαβε ο Tim Berners-Lee, θα περιέχει πάνω από 1100 σύνολα δεδομένων και η δοκιμαστική έκδοση μόλις δημοσιεύτηκε (Ιανουάριος 2010). [18]

21



Εικόνα 3.10 : Data.gov.uk (Πηγή : <http://data.gov.uk/>)

²⁰ <http://www.talis.com/platform/>

Η πρώτη online εφαρμογή Συνδεδεμένων Κυβερνητικών Δεδομένων στη Μεγάλη Βρετανία είναι η μια εθνική υπηρεσία παροχής χαρτών, η Ordnance Survey²¹, η οποία παρέχει ακριβή και ενημερωμένα γεωγραφικά δεδομένα τα οποία της τα εμπιστεύεται η κυβέρνηση, οι επιχειρήσεις και τα άτομα. Τα δεδομένα αυτά αφορούν τις διοικητικές και εκλογικές περιοχές στη Μεγάλη Βρετανία και δημοσιεύονται στην υπηρεσία αυτή και ως Συνδεδεμένα Δεδομένα χρησιμοποιώντας γνωστά λεξιλόγια (Spatial Relations Ontology, FOAF, OWL ,κλπ).[19]

Στην εικόνα που ακολουθεί φαίνεται η περιγραφή ενός πόρου, όπως δημοσιεύεται από την υπηρεσία Ordnance Survey.

Ordnance Survey – Great Britain's national mapping agency

home business leisure map shop explore magazine education about us data

Portswood

A description of the resource identified by <http://data.ordnancesurvey.co.uk/id/7000000000017711>

Preferred label	Portswood
Label	Portswood
Name	Portswood
Type	Unitary Authority Ward
Etag	"97aee58b-4200-4db8-a517-1299ae3c0dd6"
Area	280.731
Area code	UTW
Census code	00MSNC
Unit id	17711
Borders	Bitterne Park Swaythling Bassett Shirley Bevois
Partially overlaps	Southampton, Test Southampton, Itchen

The data for this description was obtained from the SPARQL service at <http://api.talis.com/stores/ordnanc> also available as: [RDF/XML](#), [Turtle](#) and [JSON](#). A free text search service is available at <http://api.talis.com>

Εικόνα 3.11 : Data Ordnance Survey

²¹ <http://data.ordnancesurvey.co.uk/>

ΚΕΦΑΛΑΙΟ 4 - Οπτικοποίηση Δεδομένων

4.1 Τι είναι η Οπτικοποίηση Δεδομένων ;

Ορισμός : «Οπτικοποίηση δεδομένων είναι η τεχνική με την οποία τα δεδομένα αναπαρίστανται σε σχηματική μορφή, με απώτερο σκοπό την ποιοτική κατανόηση των πληροφοριών που εμπεριέχουν.»

Ο όρος οπτικοποίηση δεδομένων αναφέρεται στην αναπαράσταση δεδομένων χρησιμοποιώντας γραφικά, κίνηση, τρισδιάστατες απεικονίσεις και άλλα πολυμεσικά εργαλεία. Ο σκοπός της διαδικασίας, εξαρτάται από τη χρήση για την οποία προορίζεται η εφαρμογή. Μπορεί να είναι ένα δημιούργημα καλλιτεχνικής έκφρασης, να έχει στόχο εκπαιδευτικό, επιχειρηματικό, επιστημονικό ή να είναι ένας συνδυασμός των παραπάνω.

Απλές εφαρμογές οπτικοποίησης χρησιμοποιούνται εδώ και πολλά χρόνια από τους επιστήμονες για να παρουσιάσουν τα δεδομένα των εργασιών τους (bar charts, scatter graphs, pies κλπ.), με τη βοήθεια όμως των νέων τεχνολογιών οι αρχές της οπτικοποίησης συνδυάζονται με δυναμικές εφαρμογές και μεγάλες ποσότητες δεδομένων (data sets) για να δημιουργήσουν υπερσύγχρονες εικόνες και κίνηση. Τα δεδομένα εμφανίζονται με ένα άλλο, διαφορετικό τρόπο που συχνά αποκαλύπτει νέες διασυνδέσεις και έννοιες και εξάγονται νοήματα, που διαφορετικά μπορεί να μην γίνονταν ποτέ αντιληπτά και το σημαντικό στοιχείο δεν είναι τόσο η αποσαφήνιση των εννοιών αλλά η ανάδυση της μεταξύ τους συνάφειας.

Η οπτικοποίηση περιλαμβάνει ένα ευρύ και ολοένα αυξανόμενο πεδίο που αντανακλά με δημιουργικό τρόπο την αναπαράσταση όλων των ειδών των δεδομένων οπτικά, χωρίς κανένα περιορισμό κυριολεκτικά στο ποια πληροφορία θα μετατραπεί σε εικόνα. Ο σχεδιαστής καθορίζει

ποιο οπτικό στοιχείο (χρώμα, σχήμα, μέγεθος, κίνηση κλπ.) θα αναπαραστήσει το κάθε στοιχείο των δεδομένων. Οι εικόνες μπορεί να είναι 2D, 3D, να είναι στατικές ή δυναμικές ή να επιτρέπουν αλληλεπίδραση με το χρήστη.

Ανάλογα με το περιεχόμενο για το οποίο δημιουργήθηκαν οι οπτικοποιήσεις μπορεί να αναδεικνύουν σχέσεις μεταξύ δεδομένων, να συγκρίνουν αξίες και τιμές, να ανιχνεύουν την άνοδο ή την πτώση μέσα σε χρονικές περιόδους, να βλέπουν τα μέρη ενός συνόλου, να αναλύουν ένα κείμενο από τις λέξεις του ή να βλέπουν τον κόσμο.

4.2 Ιστορική Αναδρομή

Αν και η έννοια της οπτικοποίησης δεδομένων είναι σχετικά πρόσφατη, οι ρίζες της βρίσκονται αρκετά χρόνια πίσω. Συγκεκριμένα οι πρώτοι χάρτες, που αποτελούν ένα τρόπο αναπαράστασης της πληροφορίας, δημιουργήθηκαν το μακρινό 1600 μ.Χ. Κύριος εμπνευστής της σύγχρονης οπτικοποίησης δεδομένων θεωρείται ο William Playfair (http://en.wikipedia.org/wiki/William_Playfair), Σκωτσέζος στατιστικολόγος, ο οποίος το 1801 δημιούργησε αυτό που γενικά αναφέρεται ως το πρώτο γράφημα της ιστορίας. Την σκυτάλη πήραν μετέπειτα αρκετοί ερευνητές που απεικόνισαν από γεωλογικούς χάρτες μέχρι τα θύματα πολέμων (περισσότερα εδώ <http://datavis.ca/milestones/>). Από το 1900 και μετά τα γραφήματα απέκτησαν χρώμα, εδραιώθηκαν οι κλίμακες και γενικώς άρχισαν να βελτιώνουν το αισθητικό κομμάτι τους.

Το γεγονός βέβαια που βοήθησε την εξάπλωση της οπτικοποίησης δεδομένων δεν ήταν άλλο από την εμφάνιση και εξέλιξη των ηλεκτρονικών υπολογιστών. Με την βοήθεια τους όχι μόνο μπορούμε να επεξεργαστούμε τεράστια ποσά δεδομένων αλλά και συνεχώς παράγονται καινούργια εργαλεία οπτικοποίησης. Η ιστορική εξέλιξη του μεγέθους των δεδομένων αλλά και των τρόπων οπτικοποίησης τους παρουσιάζεται στον παρακάτω πίνακα.

Πίνακας 1.3

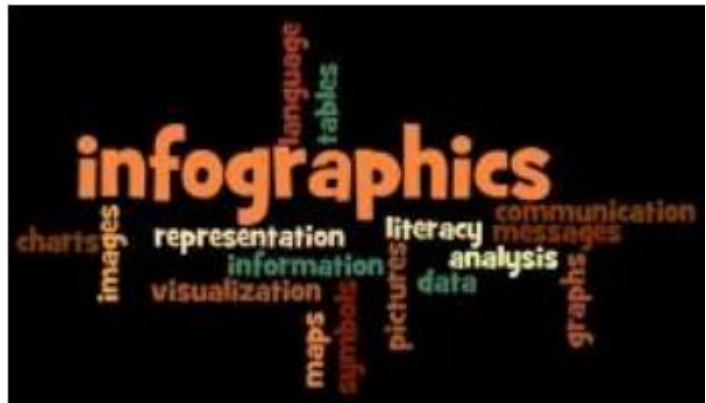
Χρονολογία	Τρόπος αποθήκευσης Δεδομένων	Μέσα επεξεργασίας Οπτικοποίησης	Πληροφορίες
1980s	Flat αρχεία	<i>Τα ίδια τα δεδομένα χρησιμοποιούνται για την εξαγωγή συμπερασμάτων</i>	http://en.wikipedia.org/wiki/Flat_file_database
1990s	Υπολογιστικά φύλλα	<i>Microsoft Excel</i>	http://en.wikipedia.org/wiki/Spreadsheet http://office.microsoft.com/el-gr/excel/ Όμικρον
1990s	Βάσεις δεδομένων	<i>Access, xBase</i>	http://en.wikipedia.org/wiki/Database http://office.microsoft.com/el-gr/access/ http://en.wikipedia.org/wiki/XBase
1990s	Σχεσιακές βάσεις δεδομένων	<i>SQL, Oracle database</i>	http://en.wikipedia.org/wiki/Relation_(database) http://www.sql.org/ http://www.oracle.com/technetwork/database/enterprise-edition/downloads/index.html
2000s	Πολυδιάστατοι κύβοι	<i>TM1</i>	http://en.wikipedia.org/wiki/OLAP_cube http://www-01.ibm.com/software/analytics/cognos/products/tm1/
2000s	<i>Πίνακες Pivot, Βάσεις δεδομένων</i>	<i>PowerPivot, Vertica</i>	http://en.wikipedia.org/wiki/Column-oriented_database

	<i>προσανατολισμένες σε στήλες</i>		http://www.microsoft.com/en-us/bi/powerpivot.aspx http://en.wikipedia.org/wiki/Pivot Table http://www.vertica.com/
2000s	Data centers	<i>VMware, Virtual Box</i>	http://en.wikipedia.org/wiki/Data_center https://www.virtualbox.org/ http://www.vmware.com/

Τέλος, το διαδίκτυο έχει συμβάλει και αυτό με την σειρά του αρκετά στην εξέλιξη της οπτικοποίησης δεδομένων. Πλέον τα δεδομένα, αλλά και οι τρόποι οπτικοποίησης μπορούν να μεταδίδονται πιο εύκολα. Η πρόσβαση σε αυτά γίνεται πιο άμεση. Το διαδίκτυο έχει προσφέρει πολλές δυνατότητες και μέσω αυτού συνεχώς προωθείται ο τομέας της οπτικοποίησης δεδομένων. Το διαδίκτυο έφερε στην επιφάνεια πολλές τεχνικές οπτικοποίησης δεδομένων, όπως τα σύννεφα ετικετών (http://en.wikipedia.org/wiki/Tag_cloud). Αυτήν την στιγμή η πιο διαδεδομένη τεχνική στο διαδίκτυο είναι τα γραφήματα πληροφορίας.

4.3 Γραφήματα Πληροφοριών (infographics)

Τα γραφήματα πληροφορίας αποτελούν κατά κάποιο τρόπο τον «πρόγονο» της οπτικοποίησης δεδομένων. Στην σημερινή εποχή έχουν κατακτήσει το διαδίκτυο. Είναι απίθανο ένας χρήστης



του διαδικτύου να μην έχει συναντήσει κάποιο. Επίσης το ίδιο απίθανο είναι κάποιος να μην έχει σχεδιάσει ένα από αυτά, ακόμα και πρόχειρο. Το έχει κάνει και απλώς δεν το κατάλαβε, αν σκεφτούμε ότι ακόμα και οι πρόχειρες οδηγίες σε χαρτί για την μετάβαση σε κάποιο μέρος είναι ένα γράφημα πληροφορίας. Η μεγάλη αύξηση της δημοτικότητας τους στο διαδίκτυο οφείλεται στο γεγονός ότι αποτελούν ένα άμεσο τρόπο παροχής πληροφοριών που γίνεται με ένα ευχάριστο οπτικό τρόπο γρήγορα. Από το πόσο άμεσα και πόσο γρήγορα γίνεται η μετάδοση της πληροφορίας κρίνεται και πόσο επιτυχημένο θεωρείται ένα γράφημα πληροφορίας. Από την σχεδίαση του, που βασίζεται κυρίως στην δημιουργικότητα του σχεδιαστή αλλά και από το πόσο καλά καταφέρνει να περάσει τις πληροφορίες που θέλει.

Υπάρχουν τρία βασικά είδη γραφημάτων πληροφορίας. Και τα τρία δημιουργούνται για τον ίδιο σκοπό, την απεικόνιση των δεδομένων, αλλά η διαφορά τους έγκειται στον τρόπο με τον οποίο αυτή επιτυγχάνεται.

- **Στατικά.** Αποτελούν το πιο διαδεδομένο είδος αυτήν την στιγμή καθώς είναι και το πιο εύκολο στην δημιουργία του. Αποτελείται από μια εικόνα που παρουσιάζει σχηματικά κάποια στοιχεία.

- **Διαδραστικά.** Τα γραφήματα αυτά μπορούν να απεικονίσουν περισσότερη πληροφορία από ότι μπορεί μια εικόνα σε μια διεπαφή. Συνήθως αποτελούνται από πολλά επίπεδα γραφημάτων. Ο χρήστης μπορεί να επιλέξει ποιο επίπεδο και κατά αντιστοιχία ποία πληροφορία θέλει να εμφανίζεται κάθε στιγμή.

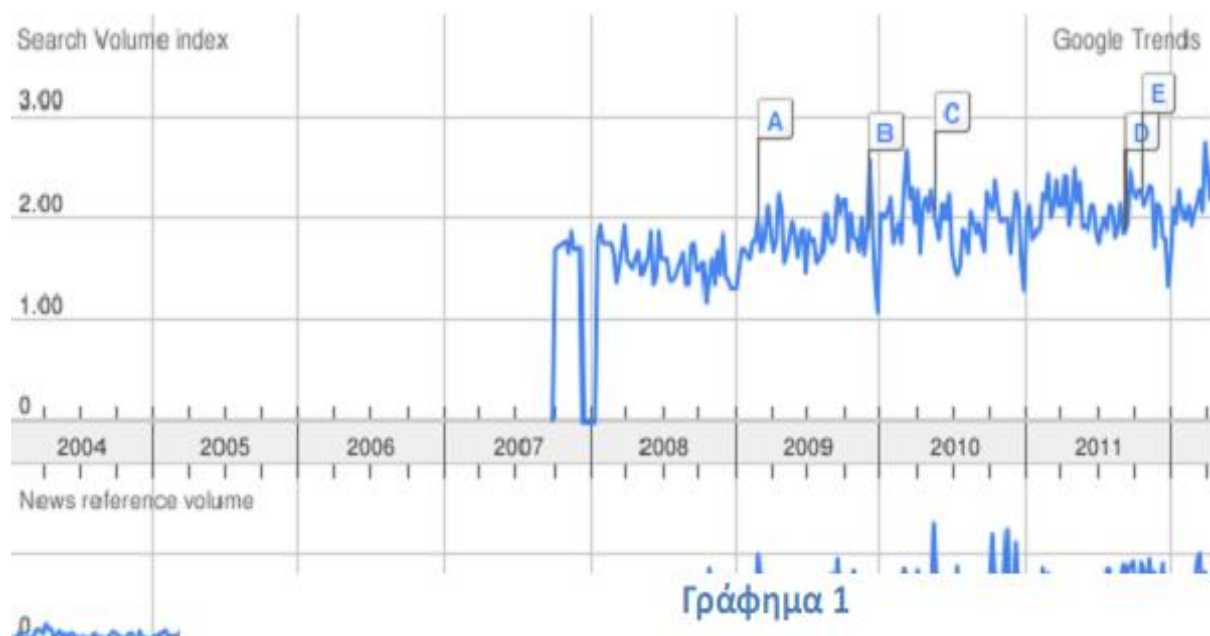
- **Με κίνηση.** Σε αυτήν την κατηγορία ανήκουν τα γραφήματα που χρησιμοποιούν κάποιο αφηγηματικό κείμενο σε συνδυασμό με πολλά εναλλασσόμενα γραφήματα με στόχο την σύνθεση κάποιου βίντεο.

Οι έννοιες οπτικοποίηση δεδομένων και γραφήματα πληροφορίας βρίσκουν εφαρμογή και οι δυο στο πεδίο του σχεδιασμού, δημιουργώντας σύγχυση στην ακριβή ερμηνεία τους. Αν και ο καθένας θα μπορούσε να τις ερμηνεύσει με τον δικό του τρόπο, οι δυο αυτές έννοιες δεν πρέπει να συγχέονται. Τα γραφήματα πληροφορίας αποτελούν ένα μόνο κομμάτι της οπτικοποίησης δεδομένων, μια τεχνική. Η οπτικοποίηση δεδομένων είναι ευρύτερη έννοια και συμπεριλαμβάνει πολλές ακόμα τεχνικές. Στα γραφήματα πληροφοριών, λόγω της μεγάλης τους εξάπλωσης, παρατηρείται μια μείωση όσο αφορά την καινοτομία. Χωρίς βέβαια αυτό να μειώνει στο ελάχιστο την αξία τους.

Το επόμενο βήμα όμως αναπόφευκτα είναι περισσότερες και διαφορετικές τεχνικές. Ωστόσο και οι δυο έννοιες έχουν κοινό παρανομαστή, που δεν είναι άλλος από την προσπάθεια εξόρυξης πληροφορίας.

4.4 Εξόρυξη Πληροφορίας

Η οπτικοποίηση δεδομένων ως τομέας έρευνας έχει αναπτυχθεί τα τελευταία χρόνια. Ωστόσο η εξέλιξη είναι ραγδαία. Αρκεί να παρατηρήσει κανείς το Γράφημα 1 (Σχήμα 4.1) που δείχνει πόσο δημοφιλή είναι η οπτικοποίηση δεδομένων, σαν έννοια, στις αναζητήσεις των χρηστών του διαδικτύου. Παλαιότερα που τα δεδομένα ήταν λιγότερα σε ποσότητα ήταν εύκολο να τα διαχειριστούμε.



Σχήμα 4.1 : Πηγή : Google trends (<http://www.google.com/trends/>)

Ωστόσο με την τεράστια αύξηση του όγκου της πληροφορίας που υπάρχει πλέον διαθέσιμη , λογικό επακόλουθο είναι να δημιουργηθεί η ανάγκη για ανάπτυξη εργαλείων που θα βοηθήσουν να αντεπεξέλθουμε σε αυτόν τον πληροφοριακό κορεσμό. Αυτό ακριβώς προσπαθεί να επιτύχει η οπτικοποίηση δεδομένων.

Η καινοτομία δεν είναι τόσο η ανακάλυψη πιο ευφάνταστων τρόπων παρουσίασης, αυτό το καταφέρνουν αρκετά καλά και τα γραφήματα πληροφορίας. Η καινοτομία βρίσκεται στην ανεύρεση νέων τρόπων συνδυασμού των δεδομένων με σκοπό την προβολή της πληροφορίας.

Αποτελεί το κατάλληλο εργαλείο για την αντιμετώπιση της «παραλυτικής ανάλυσης» (http://en.wikipedia.org/wiki/Analysis_paralysis), κατά την οποία γίνεται επανειλημμένα ανάλυση των δεδομένων, χωρίς όμως να οδηγούμαστε σε κάποιο επιθυμητό αποτέλεσμα ή απόφαση.

Ένα υπερμέγεθες σύνολο δεδομένων από μόνο του δεν προσδίδει κάποια χρησιμότητα . Είναι σχεδόν απίθανο να κατανοηθεί . Όμως με τις τεχνικές της οπτικοποίησης δεδομένων μπορεί να έρθει στην επιφάνεια όλη η πληροφορία που κρύβεται ανάμεσα τους. Η ανάδειξη αυτή πρέπει να επιτευχθεί με τρόπο αποτελεσματικό, άμεσο, ποιοτικό και ξεκάθαρο. Από αυτά τα χαρακτηριστικά κρίνεται και η επιτυχία της. Στις περισσότερες περιπτώσεις όσο μεγαλύτερο το σύνολο δεδομένων τόσο μεγαλύτερη και η χρήσιμη πληροφορία που εμπεριέχει, οπότε είναι κατανοητό πόσο σημαντικές είναι οι τεχνικές ανάλυσης του.

Με την βοήθεια της οπτικοποίησης δεδομένων ο αναγνώστης μπορεί να κατανοήσει σε βάθος τα δεδομένα. Το κυριότερο χαρακτηριστικό είναι ότι αυτό μπορεί να γίνει γρήγορα. Ο καθένας επιθυμεί να είναι ενημερωμένος. Λίγοι όμως διαθέτουν τον χρόνο για την εξέταση όλων των πληροφοριών. Η οπτικοποίηση δεδομένων δίνει την δυνατότητα στον θεατή μιας γενικής επισκόπησης. Μέσα σε λίγα δευτερόλεπτα μπορεί στοχευμένα να εξάγει την πληροφορία που επιθυμεί. Αποκόβονται όλες οι άχρηστες πληροφορίες και εστιάζει σε αυτό που πραγματικά αναζητεί και τον ενδιαφέρει. Είναι σε θέση να το εξετάσει από διαφορετικές οπτικές γωνίες. Επίσης η εξέταση αυτή επιτυγχάνεται αρκετά εύκολα. Δεν απαιτείται η απόκτηση καμίας εξειδικευμένης γνώσης για τον σκοπό αυτό. Και μπορεί να γίνει για τον καθένα ξεχωριστά σε προσωπικό επίπεδο.

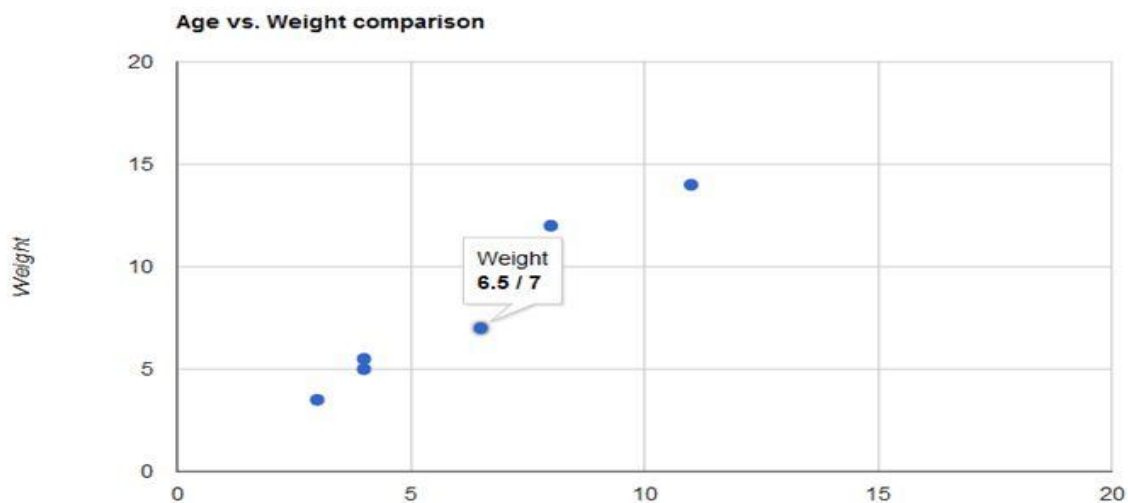
Βέβαια ο αναγνώστης δεν περιορίζεται μόνο σε συγκεκριμένες πληροφορίες αλλά μπορεί να επωφεληθεί και από τα δεδομένα στο σύνολο τους. Του παρέχεται η δυνατότητα να ανακαλύψει συσχετίσεις ανάμεσα στα δεδομένα. Συσχετίσεις που θα απαιτούσε να εξετάσει πολλά διαφορετικά σύνολα για να της ανακαλύψει. Να παρατηρήσει πολύπλοκα μοτίβα. Πολλές φορές το ζητούμενο δεν είναι κάποια σύνολα δεδομένων από μόνα τους αλλά τα αποτελέσματα που αναδεικνύονται από τις συγκρίσεις που μπορούν να γίνουν ανάμεσα τους. Ακόμα μπορεί να αναδείξει τάσεις που χωρίς την βοήθεια της οπτικοποίησης δεδομένων δεν θα ήταν σε θέση να κάνει.

Η οπτικοποίηση δεν αποτελεί μια επανάληψη των δεδομένων. Αποτελεί μια τελείως διαφορετική ερμηνεία. Παρέχει ένα εντελώς καινούργιο, συναρπαστικό τρόπο παρατήρησης τους. Τραβάει

αμέσως την προσοχή μας και καθώς επεξεργαζόμαστε τις πληροφορίες μπορεί να αλλάξει εντελώς τον τρόπο που τις βλέπαμε. Συνιστά ένα παγκόσμιο τρόπο για να κάνεις την πληροφορία προσβάσιμη στο ευρύτερο κοινό. Η «γλώσσα» της οπτικοποίησης δεδομένων είναι παγκόσμια και απολύτως κατανοητή από το σύνολο των αναγνωστών. Άλλωστε το γνωμικό «Μια εικόνα, χίλιες λέξεις» υπάρχει εδώ και χρόνια, τώρα όμως αποκτά υπόσταση και αξιοποιείται στο μέγιστο.

4.5 Οπτικοποίηση Ανοικτών Δεδομένων

Για να προχωρήσει κάποιος στην οπτικοποίηση δεδομένων πρέπει να έχει και το κατάλληλο υλικό, το οποίο μπορεί να προέρχεται από ποικίλες πηγές. Το σίγουρο είναι ότι ένα από τα καταλληλότερα είναι τα ανοιχτά δεδομένα. Και αυτός άλλωστε είναι και ο λόγος που οι διάφοροι οργανισμοί, κυβερνήσεις, επιχειρήσεις κτλ. τα προσφέρουν δωρεάν. Για να δώσουν την δυνατότητα σε όποιον επιθυμεί να τα επεξεργαστεί, με τα κατάλληλα εργαλεία οπτικοποίησης, και να τα μετατρέψει σε πληροφορία που μπορούν όλοι να καταλάβουν εύκολα.



Εικόνα 4.2 : Οπτικοποίηση Scatter chart

Στην σημερινή εποχή έχει αλλάξει κατά πολύ ο τρόπος που δεχόμαστε την πληροφορία. Καθημερινά στο διαδίκτυο συναντάμε αρκετές τεχνικές οπτικοποίησης. Αυτό μας έχει οδηγήσει στο να είμαστε εξοικειωμένοι με αυτόν τον τρόπο ενημέρωσης και όχι μόνο να νιώθουμε άνετα αλλά και πολλές φορές να τον απαιτούμε κιόλας. Συχνά είναι προτιμότερη η αναπαράσταση σε σχηματική μορφή, με χρώμα εικόνες κτλ., παρά ένα κείμενο.

Υπάρχουν κάποια χαρακτηριστικά, που αν θέλει κάποιος να δημιουργήσει μια επιτυχημένη οπτικοποίηση είναι απαραίτητο να υιοθετήσει. Η διαφορά ανάμεσα σε ένα γράφημα που καταφέρνει να μεταβιβάσει την πληροφορία και σε ένα που αποτυγχάνει, είναι μικρή.

- Η οπτικοποίηση θα πρέπει να είναι βασισμένη στο περιεχόμενο των δεδομένων και στο μήνυμα που θέλει να περάσει. Θα πρέπει να γίνει κατάλληλη επιλογή συνόλου δεδομένων για τον σκοπό που επιθυμούμε. Αυτά από μόνα τους μετά θα οδηγήσουν στον αποτέλεσμα. Η προσπάθεια τροποποίησης τους με σκοπό την παραγωγή του επιθυμητού, το μόνο που μπορεί να επιφέρει είναι παραπληροφόρηση.

- Να αναδεικνύεται η πληροφορία που είναι πιο σημαντική για τον αναγνώστη. Δεν χρειάζεται να προσπαθεί να την παρουσιάσει όλη. Τα πιο απλά μοντέλα είναι και πιο εύκολο να κατανοηθούν. Με τα κατάλληλα οπτικά ερεθίσματα να τραβάει την προσοχή εκεί που είναι απαραίτητη.

- Να παρέχουν την δυνατότητα εκμετάλλευσης της πληροφορίας. Η οπτικοποίηση δεν αφορά μόνο την ενημέρωση. Αφορά και την μετέπειτα δράση. Να προτείνει ενέργειες, να υποδεικνύει τρόπους αξιοποίησης της.

- Θα πρέπει να ξέρουμε το κοινό που απευθύνεται. Τον τρόπο με τον οποίο ο αναγνώστης θα αλληλεπιδράσει με τα δεδομένα. Η οπτικοποίηση θα πρέπει να ξεκινάει, όταν ορίσουμε τις πραγματικές ανάγκες του χρήστη. Θα πρέπει πρώτα να αντιληφθούμε πως θα χρησιμοποιηθεί η σχεδίαση και μετά να προχωρήσουμε ανάλογα στην υλοποίηση της.

- Να είναι αρκετά απλά. Πολλά είναι υπερβολικά στην σχεδίαση τους. Αυτό κρύβει το κίνδυνο ο αναγνώστης να μπερδευτεί, και να καταλήξει να κοιτάει τα γραφήματα καθαρά σαν μια μορφή τέχνης. Βέβαια πρέπει να αποφεύγονται και σχεδιάσεις που βρίσκονται στο άλλο άκρο. Τα υπεραπλουστευμένα γραφήματα που και αυτά δεν μεταδίδουν την πληροφορία, καθώς δεν κεντρίζουν το ενδιαφέρον. Είναι αναγκαίο να βρεθεί η χρυσή τομή ανάμεσα στην λειτουργικότητα και την αισθητική. Θα πρέπει να είναι σε θέση να βοηθήσουν τον αναγνώστη να κατανοήσει την πληροφορία, ενώ αυτός τα παρατηρεί με θαυμασμό.

- Η μαθηματική ακρίβεια των δεδομένων δεν χρειάζεται να είναι απόλυτη. Απλή ακρίβεια είναι αρκετή για την ανάδειξη των τάσεων.

- Να αποφεύγετε η συσσώρευση όλων των πληροφοριών που είναι διαθέσιμες. Αυτό μπορεί να επιτευχθεί με την τοποθέτηση της πληροφορίας σε διαφορετικά επίπεδα. Να υπάρχει ένα επίπεδο με την γενική απεικόνιση της πληροφορίας και μετά ο χρήστης να επιλέγει πόσο και πού θέλει να εμβαθύνει. Προσελκύοντας έτσι πολλούς διαφορετικούς τύπους αναγνωστών.
- Θα πρέπει να έχει την δυνατότητα να σταθεί από μόνο του. Να μην χρειάζεται κάποιο είδος επεξήγησης για να αντιληφθεί ο αναγνώστης πως προσπαθεί να περάσει την πληροφορία.
- Είναι αναγκαία η συνεχής ανάδραση από τους χρήστες. Μπορεί ο σχεδιαστής να θεωρεί ότι η εξόρυξη της πληροφορίας γίνεται αρκετά εύκολα γιατί αυτός το έχει υλοποιήσει και γνωρίζει ποιος είναι ο σκοπός δημιουργίας του, ωστόσο πολλοί λίγοι χρήστες να το βρίσκουν το ίδιο εύκολο. Επιπρόσθετα διαφορετικοί χρήστες αντιλαμβάνονται και διαφορετικά την πληροφορία. Αυτό που ενδεχομένως να παρατηρήσει κάποιος, μπορεί να μην το κάνει κάποιος άλλος και το αντίθετο. Η πληροφορία θα πρέπει να είναι το ίδιο ευανάγνωστη για όλους. Από τους ειδικούς στον τομέα μέχρι και για αυτούς που ασχολούνται για πρώτη φορά.

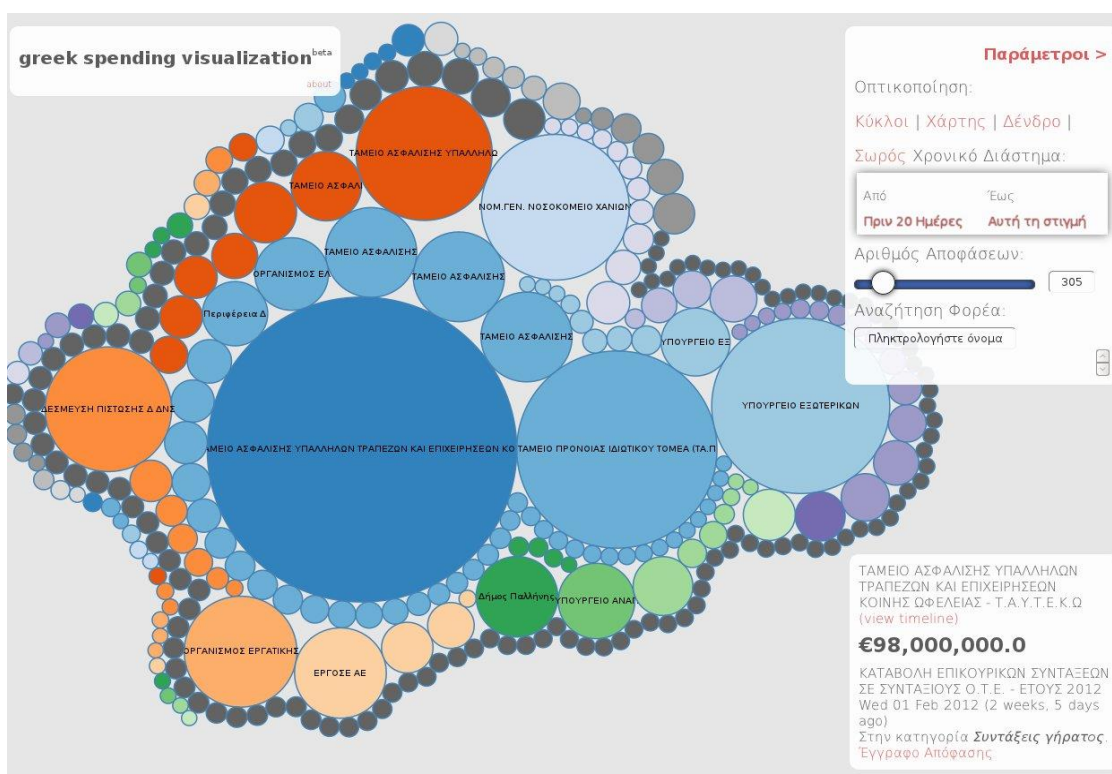
4.5.1 Περιπτώσεις Χρήσης

Τα ανοιχτά δεδομένα μπορεί να προέρχονται από πολλές διαφορετικές πηγές. Το ίδιο ποικιλόμορφο μπορεί να είναι και το περιεχόμενό τους. Εμείς θα ασχοληθούμε με αυτά που προέρχονται από τις κυβερνήσεις και σαν στόχο έχουν την προαγωγή της συμμετοχικής δημοκρατίας. Με βάση που αναλύσαμε, θα παραθέσουμε κάποια παραδείγματα οπτικοποίησης πάνω σε αυτά.

Τα παραδείγματα είναι αρκετά οπότε κατά αντιστοιχία με την παρουσίαση των συνόλων δεδομένων, διαλέξαμε να προβάλουμε τα πιο ενδιαφέροντα τόσο για Ελλάδα όσο και για το εξωτερικό.

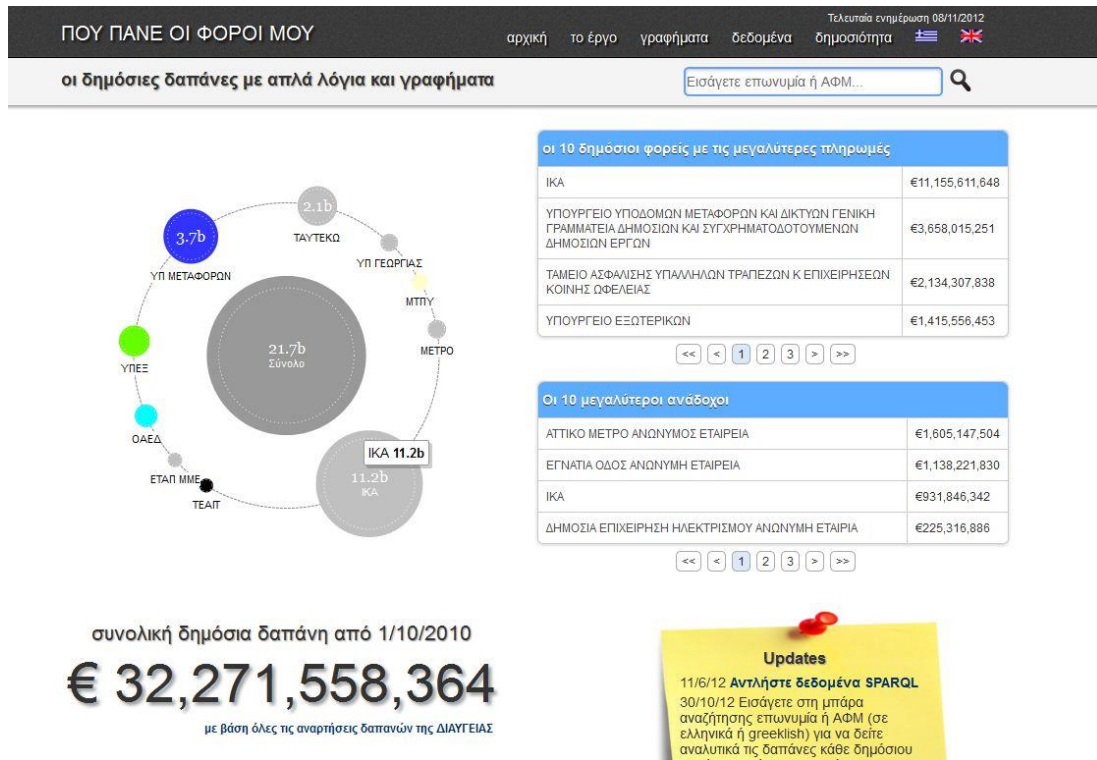
- Για την Ελλάδα :

- Greek spending visualization** (<http://greekspending.com/>) : η εφαρμογή στοχεύει στην οπτικοποίηση των δεδομένων που αφορούν τις δαπάνες του Ελληνικού Κράτους κάνοντας χρήση του προγράμματος Διαύγεια. Χάρη στην ανοικτή υλοποίηση του προγράμματος Διαύγεια, η εφαρμογή οπτικοποιεί τις 400 μεγαλύτερες σε κόστος αποφάσεις του Ελληνικού Κράτους, τις τελευταίες 20 ημέρες. Κάθε χρώμα, είναι και διαφορετική κατηγορία δαπάνης. Επιπλέον ο χρήστης μπορεί να πειραματιστεί τοποθετώντας άλλες παραμέτρους και χρονικό εύρος στην αναζήτησή του.



Εικόνα 4.3 : Εφαρμογή «greek spending visualization»

- «ΠΟΥ ΠΑΝΕ ΟΙ ΦΟΡΟΙ ΜΟΥ» (<http://publicspending.medialab.ntua.gr/index.php#>): Μια ελεύθερη, ανοιχτή, ουδέτερη και αντικειμενική διαδικτυακή εφαρμογή η οποία στοχεύει στην αποτελεσματικότερη διερεύνηση της δημόσιας δαπάνης στην Ελλάδα, όπως αυτή καταχωρείται ακριβώς στη «ΔΙΑΥΓΕΙΑ» .Σκοπός της εφαρμογής είναι η αξιοποίηση ανοικτών δεδομένων (Open Data) που σχετίζονται με τις δημόσιες δαπάνες καθώς και η βελτίωση της διαδικασίας λήψης αποφάσεων σε ό,τι αφορά τις δημόσιες δαπάνες



Εικόνα 4.4 : Εφαρμογή «ΠΟΥ ΠΑΝΕ ΟΙ ΦΟΡΟΙ ΜΟΥ»

- **Ανοιχτά Δημόσια Δεδομένα** (<http://projects.thodoris.net>). Εφαρμογή για αναζήτηση στις αποφάσεις μόνο όμως με βάση τον φορέα. Αυτό που το κάνει διαφορετικό είναι ότι δίνει την δυνατότητα για παρακολούθηση του κρατικού προϋπολογισμού.



✚ Για το εξωτερικό :

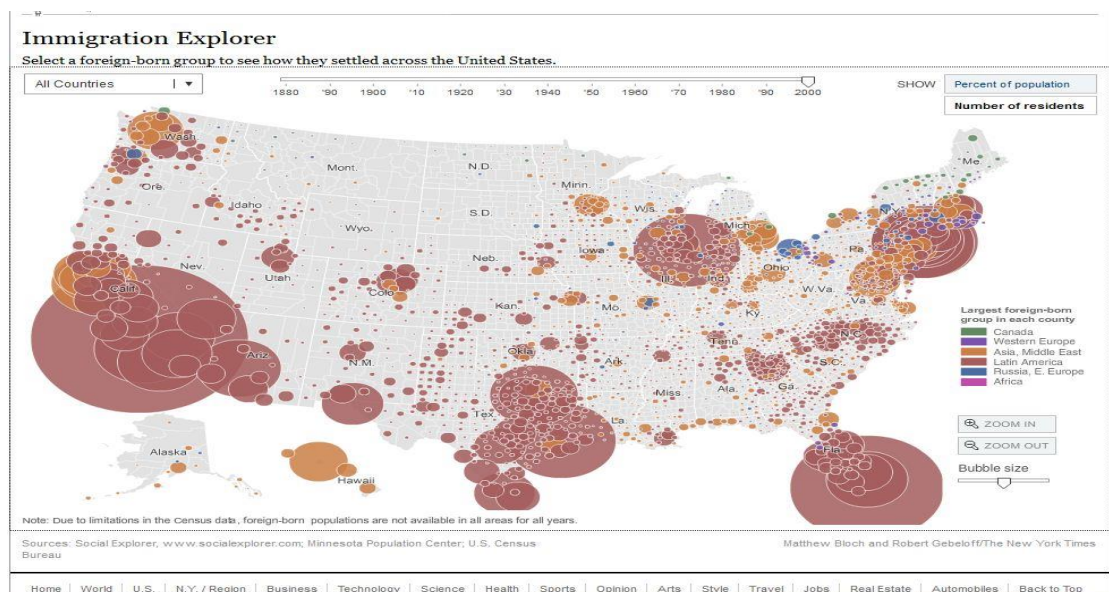
- **Recovery.gov** (<http://www.recovery.gov/Pages/default.aspx>). Απεικονίζει με ποιο τρόπο ξοδεύονται τα κυβερνητικά κεφάλαια των Ηνωμένων Πολιτειών Αμερικής. Παρέχει επίσης και κάποια μορφή ελέγχου με την δυνατότητα που προσφέρει στους πολίτες αναφοράς περιπτώσεων απάτης, σπατάλης και εκμετάλλευσης. Παρόμοιες προσπάθειες αποτελούν το Look at Cook (<http://lookatcook.com/>) για τα κρατικά κεφάλαια του Illinois, το Texas recovery (<http://goo.gl/gQmRd>) για τα αντίστοιχα του Texas και το Kenya Budget (<http://goo.gl/8RZol>) για αυτά της Κένυας.



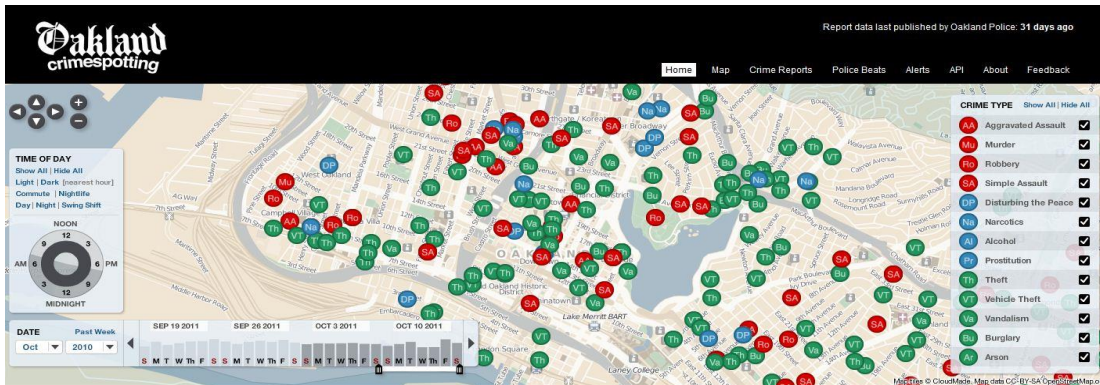
- **Where does my money go?** (<http://wheredoesmymoneygo.org/>). Παρουσιάζει πως χρησιμοποιούνται τα χρήματα που δίνουν οι πολίτες του Ηνωμένου Βασιλείου μέσω της φορολογίας. Εμφανίζει αναλυτικά τα ποσά που παραχωρούνται σε κάθε τομέα (υγεία, άμυνα κτλ.).



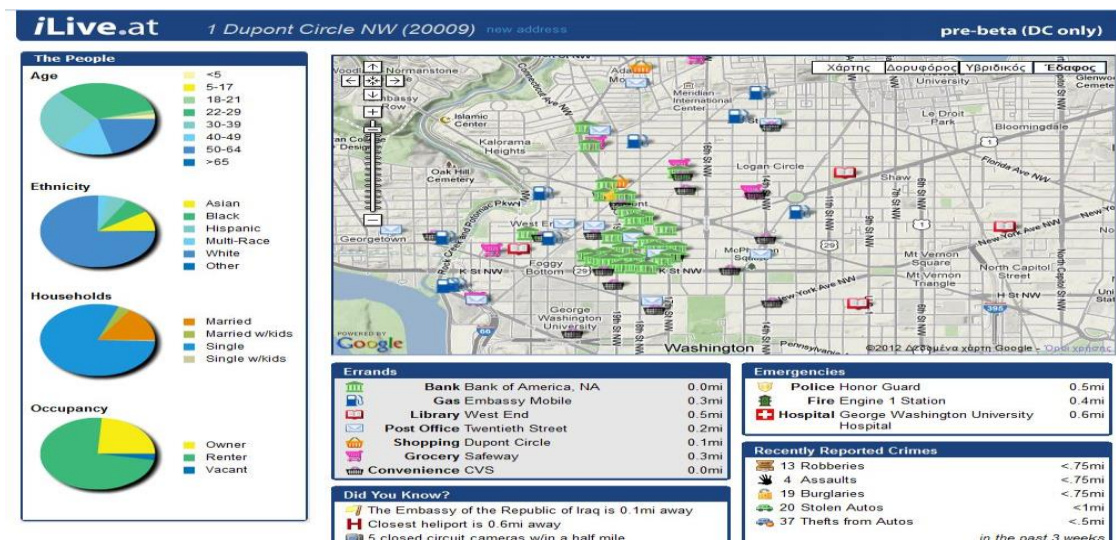
- **Immigration Explorer** (<http://goo.gl/f85h>). Εκθέτει τον αριθμό, τόσο σε ποσοστό όσο και πραγματικούς αριθμούς, των μεταναστών σε κάθε περιοχή των Ηνωμένων Πολιτειών Αμερικής καθώς και την περιοχή από την οποία μετανάστευσαν. Υπάρχει η δυνατότητα επιλογής συγκεκριμένης δεκαετίας (1880-2000) για την οποία θα γίνει η αναπαράσταση.



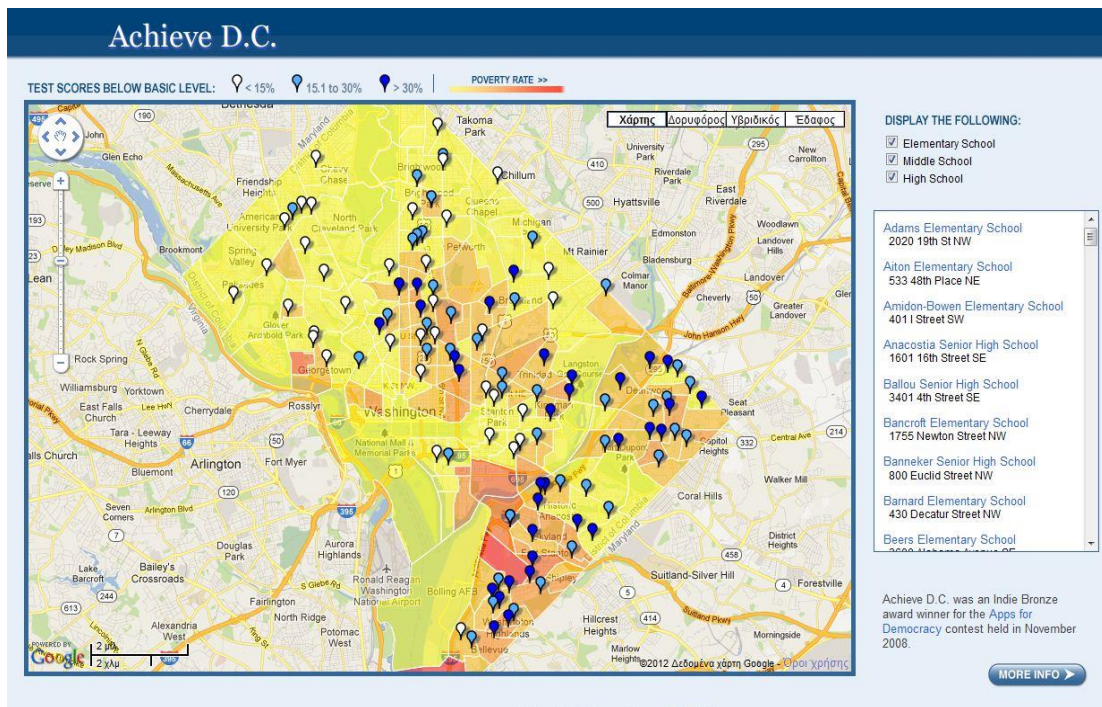
- Oakland Crimespotting** (<http://goo.gl/IQBZQ>). Παρουσιάζει τα εγκλήματα που διεπράχθησαν στην πόλη του Oakland. Το είδος τους και την ακριβή τοποθεσία. Παρέχει την δυνατότητα απεικόνισης αποτελεσμάτων σε συγκεκριμένο ωρολογιακό παράθυρο αλλά και ημερομηνία. Παρόμοια προσπάθεια, αλλά αρκετά υποδεέστερη, αποτελεί το crime.Stats (<http://goo.gl/WgAOC>), το οποίο βασίζεται στα κυβερνητικά δεδομένα του Ηνωμένου Βασιλείου και καταδεικνύει την εγκληματικότητα στην περιοχή επιλογής του χρήστη.



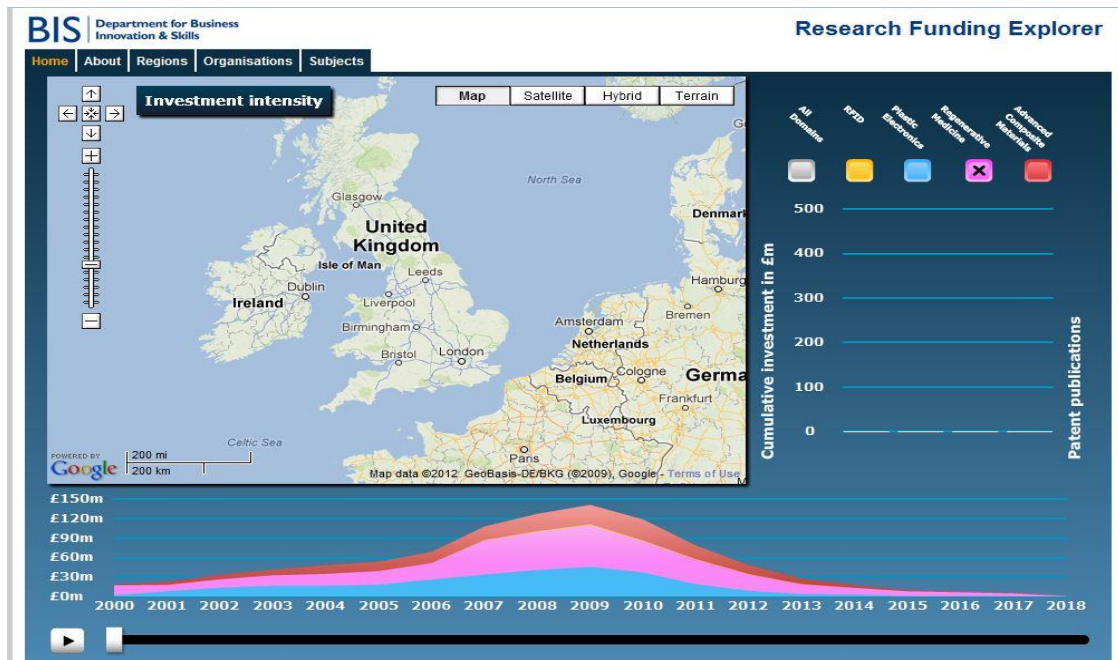
- iLiveat** (<http://www.ilive.at/>). Προσφέρει πληροφορίες για την γειτονιά που επιθυμείς, μόνο στην πόλη της Washington προς το παρών. Αργότερα θα εξελιχθεί για όλες τις περιοχές των Ηνωμένων Πολιτειών Αμερικής. Πληροφορίες για τους κατοίκους όπως οικογενειακή κατάσταση, εθνικότητα, ηλικία. Αλλά και για υπηρεσίες όπως το κοντινότερο νοσοκομείο, τράπεζα, ταχυδρομείο και αρκετές ακόμα.



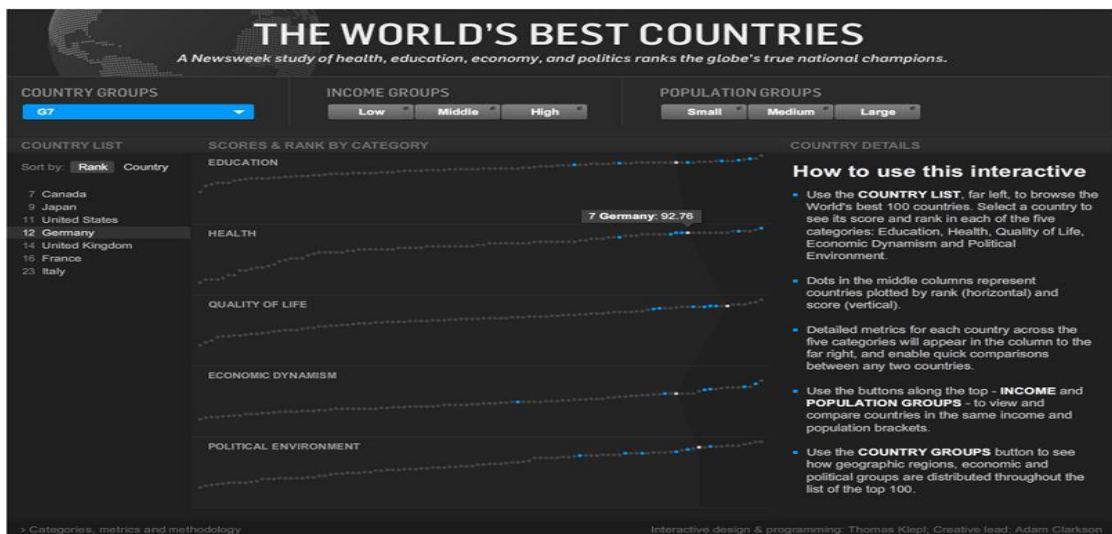
- **Achieve D.C.** (<http://achievedc.mapkist.com/>). Απεικονίζει την αξιολόγηση κάθε σχολείου στην πόλη της Washington σε αντιδιαστολή με τον συντελεστή φτώχειας της περιοχής που εδρεύει.



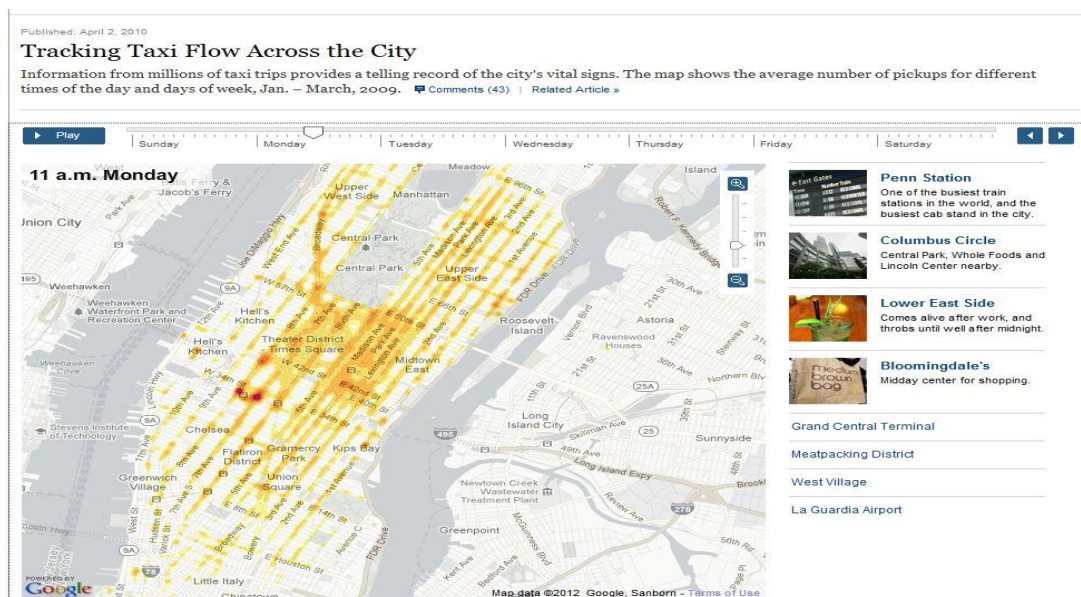
- **BIS** (<http://bis.clients.talis.com/>). Παραθέτει τα ποσά των επενδύσεων, και σε ποιους τομείς, που έχει κάνει η κυβέρνηση του Ηνωμένου Βασιλείου.



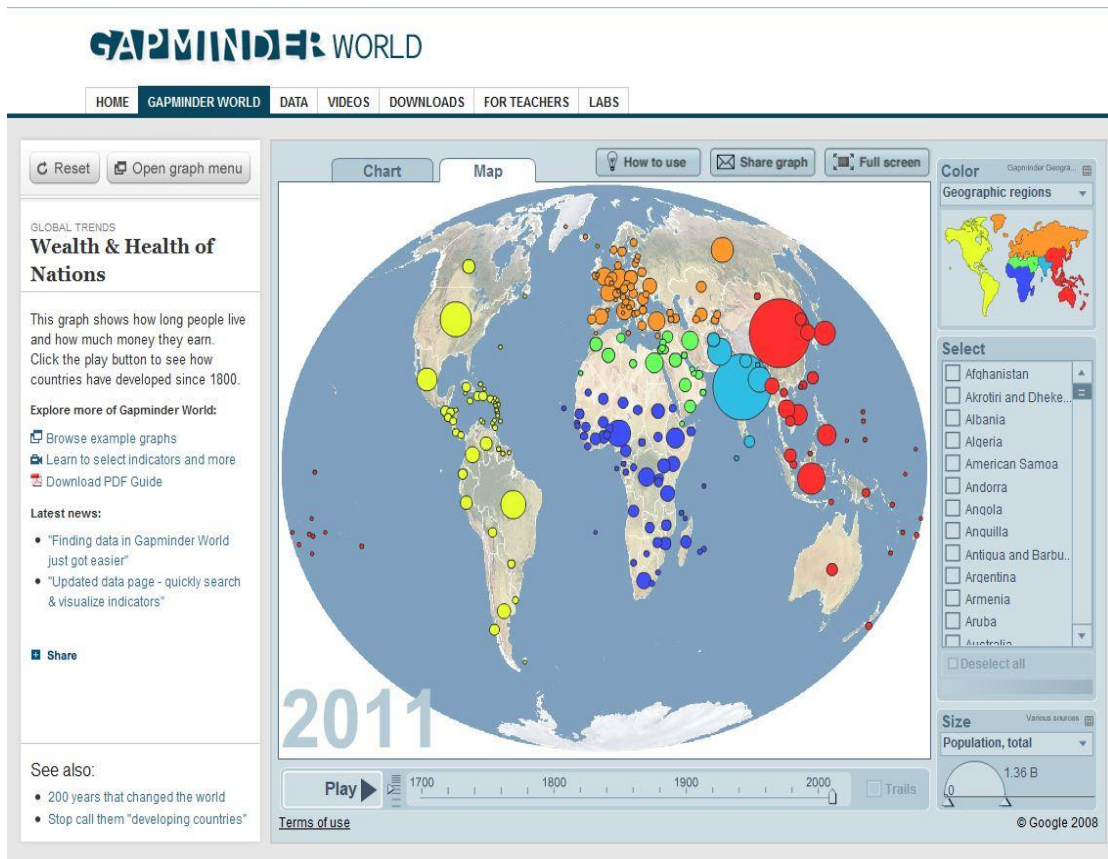
- **The world's best countries** (<http://goo.gl/FZb2Z>). Παρουσιάζει τις 100 καλύτερες χώρες, ξεχωριστά στους τομείς της υγείας, της οικονομίας και της εκπαίδευσης. Όπως επίσης και αναλυτικά στατιστικά για κάθε μία από αυτές.



- **Cab flow** (<http://goo.gl/4wpl>). Παρουσιάζει τις διαδρομές των ταξί κατά την διάρκεια της ημέρας, την τελευταία εβδομάδα στην πόλη της Νέας Υόρκης. Παρόμοιας φιλοσοφίας είναι και το **London Underground** (<http://traintimes.org.uk/map/tube/>) το οποίο δείχνει σε πραγματικό χρόνο την θέση όλων των τρένων του υπογείου του Λονδίνου. Καθώς και το **train map** (<http://www.swisstrains.ch/>) που είναι το ίδιο αλλά για τα τρένα των σιδηροδρομικών γραμμών της Ελβετίας.



- **Gapminder World** (<http://www.gapminder.org/>). Είναι μια εφαρμογή που οι χώρες του κόσμου εμφανίζονται σαν φούσκες και ο πληθυσμός τους αντιστοιχεί στο μέγεθος της φούσκας. Υπάρχει η δυνατότητα να συνδυάσει κανείς τους άξονες Χ και Υ για να συγκρίνει το προσδόκιμο ζωής με το κατά κεφαλή εισόδημα και το εργαλείο μας δείχνει πως κινείται η φούσκα της κάθε χώρας στο χρόνο.



4.5.2 Εργαλεία Οπτικοποίησης Δεδομένων

Συνέχεια εξελίσσονται καινούργια εργαλεία οπτικοποίησης δεδομένων. Η εργασία των σχεδιαστών γίνεται ακόμα πιο εύκολη. Πολύπλοκες αναλύσεις γίνονται εφικτές. Αυτό θα προσελκύσει ακόμα περισσότερους να επεξεργαστούν τα δεδομένα και να οπτικοποιήσουν τα αποτελέσματα. Και όλα αυτά λειτουργούν υπέρ αυτών που θα καταναλώσουν στο τελικό στάδιο την πληροφορία.

Παρακάτω παρουσιάζεται μια λίστα με τα πιο διαδομένα εργαλεία.

- **Google charts** (<https://developers.google.com/chart/>). Το Google chart API, είναι μια νέα υπηρεσία της Google για προγραμματιστές, μας επιτρέπει να δημιουργούμε δυναμικά διαγράμματα (charts) χωρίς τη χρήση κάποιου dll ή άλλου component και χωρίς να καταναλώνουμε την υπολογιστική ισχύ του server μας.
- **Impure** (<http://www.impure.com/>). Μια εφαρμογή διαδικτύου που βοηθάει στην οπτικοποίηση δεδομένων που προέρχονται από οποιαδήποτε πηγή και στην μετέπειτα δημοσίευση τους.
- **Many Eyes** (<http://www-958.ibm.com/software/data/cognos/maneyeyes/>). Ανεβάζεις τα δεδομένα σου, παρέχει αρκετές μεθόδους οπτικοποίησης και στην συνέχεια τα κοινοποιείς.
- **Tableau Public** (<http://www.tableausoftware.com/public>). Βοηθάει στην σχεδίαση προτείνοντας διάφορους τρόπους οπτικοποίησης των δεδομένων σου. Το τελικό αποτέλεσμα μπορείς να το ενσωματώσεις σε οποιαδήποτε ιστοσελίδα ή να το δημοσιεύσεις.
- **R Project** (<http://www.r-project.org/>). Μια καθαρά προγραμματιστική γλώσσα η οποία χρησιμοποιείται τόσο στην ανάλυση όσο και στην οπτικοποίηση των δεδομένων.
- **High Charts** (<http://www.highcharts.com/>). Αποτελεί μια έτοιμη βιβλιοθήκη οπτικοποίησης δεδομένων, βασισμένη στην γλώσσα προγραμματισμού Javascript (<http://en.wikipedia.org/wiki/JavaScript>). Παρόμοιες βιβλιοθήκες αποτελούν τα Exhibit (<http://simile-widgets.org/exhibit/>), infoVis (<http://thejit.org/>), protoVis (<http://mbostock.github.com/protovis/>) και το D3 (<http://d3js.org/>).
- **Zoho Reports** (<https://reports.zoho.com>). Προσφέρει την δυνατότητα για την δημιουργία γραφημάτων κατευθείαν από βάσεις δεδομένων ή από υπολογιστικά φύλλα.
- **gnuplot** (<http://www.gnuplot.info/>). Παρέχει την δυνατότητα για τρισδιάστατα γραφήματα. Παρόμοιες δυνατότητες έχουν και τα Scigraphica (<http://scigraphica.sourceforge.net/>), paraview (<http://www.paraview.org/>) και VTK (<http://www.vtk.org/>).

4.6 Κριτική Ανάλυση

Μέσα από ένα γράφημα ο αναγνώστης μπορεί να κατανοήσει αμέσως όλα αυτά, που αλλιώς θα χρειαζόνταν ένα τεράστιο αριθμό από νούμερα και επεξηγήσεις. Επίσης πλέον του παρέχεται η δυνατότητα να εντυφώσει στα δεδομένα όσο αυτός επιθυμεί. Για να επιτευχθούν όμως τα αναμενόμενα αποτελέσματα, που αναλύσαμε παραπάνω, θα πρέπει να χρησιμοποιούνται τα κατάλληλα σχήματα και χρώματα για αυτόν τον σκοπό. Ακόμα ο σχεδιαστής θα πρέπει να επιλέξει με προσοχή όλες τις παραμέτρους του γραφήματος. Το μέγεθος, την διάταξη, την χωρική θέση. Όλα αυτά παίζουν σημαντικό ρόλο. Είναι αναγκαίο να αποφευχθούν τα λεγόμενα «γραφήματα σκουπιδιών» (<http://en.wikipedia.org/wiki/Chartjunk>). Γραφήματα που αποσπάνε την προσοχή του χρήστη αντί να τον πληροφορούν.

Η σύγχυση δεν είναι έννοια που συμβαδίζει με την προσπάθεια ανάγνωσης της πληροφορίας. Προέρχεται από την κακή απεικόνιση της. Ο αναγνώστης όταν κοιτάζει το γράφημα δεν θα πρέπει να αναρωτιέται που να δώσει προσοχή. Τι είναι αυτό που απεικονίζει ή πως θα πρέπει να χρησιμοποιήσει την πληροφορία. Όλα αυτά επιβάλλεται να είναι ξεκάθαρα με την πρώτη ματιά.

Αν και η οπτικοποίηση δεδομένων έχει αναπτυχθεί πολύ, απέχει ακόμα αρκετά από την τελική της μορφή και τον σκοπό για τον οποίο εξελίχθηκε. Μέχρι κάθε χρήστης του διαδικτύου να χρησιμοποιεί πληροφορίες που είναι αποτέλεσμα οπτικοποίησης κάποιων πολύπλοκων δεδομένων δεν έχει ολοκληρωθεί. Ωστόσο έχει γίνει μια πολλά υποσχόμενη αρχή.

Πλεονεκτήματα :

Η οπτικοποίηση είναι ένας τρόπος να τιθασεύσουμε πληροφορίες, να ανακαλύψουμε τάσεις και διασυνδέσεις που διαφορετικά θα παρέμεναν αόρατες. Φέρνει στην επιφάνεια θέματα, ιδέες και στοιχεία πρωτότυπα, με αποτέλεσμα την κατανόηση και την πληρέστερη επεξεργασία τεραστίων ποσοτήτων πληροφορίας. Η προσέγγιση αυτή είναι πολύτιμη στην οικοδόμηση της γνώσης. Ενισχύει την δημιουργική σκέψη και θέτει κρίσιμα ερωτήματα. Οι δεξιότητες που απαιτούνται για τη χρήση και την κατανόηση της διαδικασίας είναι σημαντικές.

Μειονεκτήματα :

Αν τα στοιχεία στα οποία βασίζεται η οπτικοποίηση δεν είναι αξιόπιστα η πληροφορία που βγαίνει είναι προβληματική ακόμη κι αν τα δεδομένα είναι αξιόπιστα μια φτωχή οπτικοποίηση μπορεί να μη δείξει τίποτε ενδιαφέρον, η να υπερβάλλει στη σπουδαιότητα κάποιων τάσεων οδηγώντας σε εσφαλμένα συμπεράσματα. Οι περισσότεροι άνθρωποι επίσης δεν είναι ακόμη εξοικειωμένοι με τις οπτικές αναπαραστάσεις για να μπορέσουν να εκμεταλλευτούν τις δυνατότητές τους.

ΚΕΦΑΛΑΙΟ 5 - Η Εφαρμογή (<http://greekopendata.host22.com/>)

5.1. Στόχος

Τα *Ανοιχτά Δεδομένα* και ειδικότερα τα *Ανοιχτά Κυβερνητικά Δεδομένα*, είναι μια τεράστια πηγή δεδομένων που δεν έχει ακόμα αξιοποιηθεί. Πολλά μεμονωμένα άτομα, αλλά και οργανισμοί, συλλέγουν ένα ευρύ φάσμα διαφορετικών τύπων δεδομένων, προκειμένου να εκπληρώσουν τις εργασίες τους. Τα κυβερνητικά δεδομένα είναι ιδιαίτερα σημαντικά από αυτήν την άποψη, τόσο λόγω της ποσότητας και της καθολικότητάς τους, όσο και λόγω του ότι αυτά τα δεδομένα είναι, από νομικής άποψης, δημόσια δεδομένα και συνεπώς θα μπορούσαν να είναι ανοιχτά και διαθέσιμα σε άλλους να τα χρησιμοποιήσουν.

Ο Νόμος 3979/2011 (Νόμος για την Ηλεκτρονική Διακυβέρνηση) προβλέπει την ελεύθερη διάθεση των δεδομένων χωρίς ουσιαστικούς περιορισμούς, με τρόπο δηλαδή που επιτρέπει την περαιτέρω χρήση και επεξεργασία τους από τον καθένα. Όμως αυτό που κρίνει τελικά την ουσία του δικαιώματος δεν είναι μόνο το νομικό πλαίσιο, αλλά επίσης, οι οργανωτικές και τεχνολογικές επιλογές που το υλοποιούν. Στο τομέα αυτό απαιτούνται περαιτέρω βήματα και συντονισμός καθώς η Ελλάδα υστερεί σε σχέση με άλλες χώρες.

Υπάρχουν πολλοί τομείς που μπορούμε να περιμένουμε πως τα *Ανοιχτά Δεδομένα* μπορούν να έχουν αξία και υπάρχουν ήδη παραδείγματα που αναδεικνύουν πώς αυτά έχουν χρησιμοποιηθεί. Υπάρχουν, επίσης, πολλές ομάδες ανθρώπων και οργανισμών που μπορούν να επωφεληθούν από τη διαθεσιμότητα των ανοιχτών δεδομένων, συμπεριλαμβανομένων των ίδιων των κυβερνήσεων. Την ίδια στιγμή, είναι αδυνατόν να προβλέψουμε με ακρίβεια πώς και πού μπορούν να προσδώσουν αξία μελλοντικά. Σε αυτήν την εφαρμογή γίνεται μια προσπάθεια για την αξιοποίηση Δημόσιων Ανοιχτών Δεδομένων τόσο με στατικό, όσο και με δυναμικό τρόπο, έτσι ώστε να αυξηθεί η πληροφοριακή τους αξία. Οι πηγές, από τις οποίες συλλέχθηκαν τα δεδομένα είναι το Διαύγεια (μέσω των apis του Διαύγεια), το ΣΕΤΕ, Η Πυροσβεστική και η Αστυνομία.

5.2. Προδιαγραφές

Η εφαρμογή δημιουργήθηκε βάσει κάποιων προδιαγραφών που είχαν σαν κύριο σκοπό την επίτευξη των στόχων που αναλύθηκαν (βλ. Ενότητα 5.1). Οι βασικότερες προδιαγραφές είναι :

- Να παρέχεται μια γρήγορη, άμεση εποπτική ματιά σε δαπάνες από κυβερνητικούς φορείς καθώς και άλλα στοιχεία.
- Να αντλούνται εύκολα πληροφορίες.
- Να αναδεικνύονται οι τάσεις.
- Να είναι εύχρηστο.
- Να γίνονται εύκολα συγκρίσεις.

5.3. Διαύγεια

Στα πλαίσια των ανοιχτών δεδομένων (βλ. Κεφάλαιο 2) αλλά και της συμμετοχικής δημοκρατίας το ελληνικό κράτος έχει προχωρήσει στην υλοποίηση του προγράμματος Διαύγεια (<http://diavgeia.gov.gr/>). Μέσω αυτού όλοι οι κρατικοί φορείς (πανεπιστήμια, δήμοι, νοσοκομεία κτλ.) είναι υποχρεωμένοι να αναρτούν τις αποφάσεις τους στο διαδίκτυο. Για την ακρίβεια για να εγκριθεί η οποιαδήποτε απόφαση πρέπει πρώτα αυτή να πάρει ένα μοναδικό αριθμό διαδικτυακής ανάρτησης (ΑΔΑ) και έπειτα να εκτελεστεί. Κάθε απόφαση χαρακτηρίζεται από ένα συγκεκριμένο τύπο (π.χ. νόμος, δαπάνη, διορισμός κτλ.) και εντάσσεται σε μια συγκεκριμένη θεματική ενότητα (π.χ. οικονομία, εκπαίδευση, στρατιωτική θητεία κτλ.) Ο αντικειμενικός σκοπός δημιουργίας του είναι να ενισχύσει την διαφάνεια και να επιτρέψει στους πολίτες να έχουν πρόσβαση σε όλες τις κυβερνητικές αποφάσεις.

Η δημοσίευση των αποφάσεων είναι η αρχή. Το επόμενο βήμα είναι η παροχή κατάλληλων εργαλείων για την διαχείριση τους. Και αυτό προσφέρεται μέσω του API (Application Programming Interface <http://goo.gl/ByBW>)

της Διαύγειας

(<http://opendata.diavgeia.gov.gr/>). Μέσα

από αυτό ο χρήστης μπορεί να ορίσει μια σειρά από παραμέτρους (π.χ. χρονικό παράθυρο, φορέας κτλ.) και στην συνέχεια

να προχωρήσει στην αναζήτηση των αποφάσεων. Τα αποτελέσματα-αποφάσεις που θα του επιστραφούν θα είναι μόνο αυτά που πληρούν τις προϋποθέσεις που έθεσε. Ο χρήστης μπορεί να χρησιμοποιήσει τα αποτελέσματα της αναζήτησης όπως αυτός επιθυμεί.



Το API της διαύγειας αποτέλεσε και την αφορμή για την δημιουργία της εφαρμογής μας και όχι μόνο. Αρκετές εργασίες έχουν στηριχτεί σε αυτό για την υλοποίησή τους. Χωρίς αυτό θα ήταν πάρα πολύ δύσκολο, αν όχι αδύνατο, να επεξεργαστεί κάποιος τις αποφάσεις και να παράγει ένα τελικό προϊόν, όπου και αν στοχεύει αυτό.

Όπως και σε κάθε καινούργια προσπάθεια, έτσι και εδώ, υπάρχουν θετικά και αρνητικά στοιχεία. Αυτά αποτελούν προσωπικές απόψεις που προήλθαν από την επαφή μου με το API. Ωστόσο σε αρκετά από αυτά συμφωνούν όσοι ασχολήθηκαν με αυτό. Στα θετικά μπορούμε να συμπεριλάβουμε τα:

- Η δημιουργία του προγράμματος Διαύγεια και η υλοποίηση του API που βασίστηκε πάνω του, έγιναν σε παγκόσμιο επίπεδο αρκετά γρήγορα. Η αρχή στο άνοιγμα των κυβερνητικών δεδομένων έγινε τον Μάιο του 2009 από τις Ηνωμένες Πολιτείες Αμερικής μέσω της επίσημης ιστοσελίδας της κυβέρνησης και ακολούθησε λίγο μετά, τον Σεπτέμβριο του ίδιου έτους, το Ηνωμένο Βασίλειο. Η Ελλάδα δεν άργησε να ακολουθήσει. Τοποθετώντας τις χώρες με βάση την χρονολογία που άνοιξαν τα δεδομένα τους βρίσκεται στην έκτη θέση, έχοντας προβεί στην δημοσίευση τους ένα χρόνο νωρίτερα από χώρες όπως η Γαλλία και η Ισπανία, με αυτό να γίνεται τον Νοέμβριο του 2010.

- Η ταξινόμηση των αποφάσεων γίνεται με μεγάλη επιτυχία και είναι ιδιαίτερα σαφής. Οι παράμετροι του API καλύπτουν όλα τα πεδία που τις χαρακτηρίζουν. Είναι αρκετά ξεκάθαρο και απλό στην τεκμηρίωση του. Ο χρήστης μπορεί εύκολα να εξάγει ακριβώς τις αποφάσεις που τον ενδιαφέρουν.

Στον αντίποδα τώρα υπάρχουν και οι ακόλουθες ελλείψεις:

- Δεν υπάρχει ένας καθολικός και υποχρεωτικός τρόπος εισαγωγής των αποφάσεων. Ο εκάστοτε αρμόδιος μπορεί να τις εισάγει με την δική του ερμηνεία. Για παράδειγμα κάποιος μπορεί να χρησιμοποιήσει την τελεία για να προσδιορίσει τις χιλιάδες ενός ποσού ενώ κάποιος άλλος για να προσδιορίσει το δεκαδικό μέρος αυτού. Τέτοια λάθη γίνονται σε αρκετά πεδία, αλλά στα αριθμητικά έχουν το μεγαλύτερο αντίκτυπο. Δημιουργεί προβλήματα στην προσπάθεια ανάλυσης των δεδομένων καθώς αυτή ενδεχομένως να στηρίζεται σε λάθος ερμηνευμένα στοιχεία. Είναι αναγκαίο να υπάρξει κάποιος έλεγχος εγκυρότητας ως προς την μορφή που εισάγονται τα δεδομένα. Καταχωρίσεις που δεν συμβαδίζουν με την προκαθορισμένη μορφή θα πρέπει να απορρίπτονται.
- Πολλές φορές τα δεδομένα στα πεδία που επιστρέφει το API δεν αντανακλούν τα πραγματικά στο κείμενο της απόφασης. Θα πρέπει να δημιουργηθεί κάποια μορφή επικύρωσης-επαλήθευσης.
- Αν και όπως προαναφέραμε το API δημιουργήθηκε αρκετά νωρίς σε σχέση με τις υπόλοιπες χώρες εδώ και αρκετό καιρό είναι στάσιμο σε ότι αφορά την εξέλιξη του. Παρόλο που είχε όλο αυτόν τον καιρό για να πάρει την ανάδραση που χρειάζεται από τους χρήστες. Χώρες που δεν διέθεταν τα δεδομένα τους τόσο νωρίς όσο η Ελλάδα αυτήν την στιγμή έχουν πολύ καλύτερα και εξελιγμένα συστήματα διαχείρισης τους.
- Τα δεδομένα που διατίθενται περιορίζονται ακόμα στο σύνολο των αποφάσεων των κυβερνητικών οργάνων. Θα πρέπει η Ελλάδα να παραδειγματιστεί από τις υπόλοιπες χώρες (π.χ. η Κένυα προσφέρει 515 διαφορετικά σύνολα δεδομένων) και να προχωρήσει στο άνοιγμα και άλλων δεδομένων.
- Δεν παρέχει καμία υποστήριξη στους χρήστες που προσπαθούν να δημιουργήσουν εργαλεία βασισμένα στο API. Τα υλοποιούν από μόνοι τους, για προσωπικούς λόγους ο καθένας, και έπειτα δεν γίνεται καμία προσπάθεια ανάδειξης τους από την μεριά του

κράτους. Με αυτόν τον τρόπο δεν προσφέρεται και κανένα απολύτως κίνητρο ώστε να προσελκύσει ακόμα περισσότερους ιδιώτες να δουλέψουν πάνω στα δεδομένα.

Γενικότερα η όλη υλοποίηση της Διαύγειας αφήνει στο τέλος μια αίσθηση ότι πραγματοποιήθηκε περισσότερο για να δείξει ότι η Ελλάδα συμβαδίζει με τον κλίμα της εποχής, που προστάζει το άνοιγμα των κυβερνητικών δεδομένων, και όχι για να προσφέρει κάτι ουσιαστικό. Βέβαια πολλές χώρες δεν έχουν προβεί ακόμα σε τέτοιες ενέργειες. Το πρόγραμμα Διαύγεια αποτελεί το πρώτο σημαντικό βήμα που όμως χρειάζεται αρκετά ακόμα για να φτάσει σε ικανοποιητικό επίπεδο. Ας ελπίσουμε ότι στο μέλλον θα δώσουν την απαραίτητη προσοχή που απαιτούν τέτοιες προσπάθειες ώστε να επιτύχουν τον αντικειμενικό τους σκοπό.

5.4. Υλοποίηση

5.4.1 Τεχνολογίες

Παρακάτω θα αναλύσουμε όλες τις τεχνολογίες που χρησιμοποιήθηκαν στην υλοποίηση της εφαρμογής.

Συγκεκριμένα έχουμε:

- **HTML** (<http://el.wikipedia.org/wiki/HTML>). Αποτελεί τη βασική γλώσσα προγραμματισμού για την δημιουργία των αντικειμένων μιας ιστοσελίδας.
- **CSS** (<http://www.w3.org/Style/CSS/>). Ενσωματώνεται στην HTML και παρέχει την δυνατότητα για μορφοποίηση των αντικειμένων της ιστοσελίδας.
- **Javascript** (<http://en.wikipedia.org/wiki/JavaScript>). Μια γλώσσα προγραμματισμού η οποία χρησιμοποιείται για να εκτελούνται διάφορες λειτουργίες στις ιστοσελίδες.
- **JQuery** (<http://jquery.com/>). Μια βιβλιοθήκη βασισμένη στην γλώσσα Javascript. Επιλέχθηκε επειδή παρέχει την δυνατότητα για συγγραφή πολύ λιγότερου κώδικα, για τις ίδιες λειτουργίες, όπως επίσης και γιατί κάνει την υλοποίηση κάποιων λειτουργιών (π.χ. χειρισμός γεγονότων) πολύ πιο εύκολη.
- **AJAX** (<http://www.w3schools.com/ajax/default.asp>). Μια τεχνολογία που παρέχει την δυνατότητα να ανταλλάσσει δεδομένα με τον server και να ανανεώνει συγκεκριμένα

αντικείμενα της ιστοσελίδας ασύγχρονα. Χρησιμοποιήθηκε καθώς θα έπρεπε να ανανεώνεται το κεντρικό γράφημα, ανάλογα με τις επιλογές του χρήστη, χωρίς όμως να φορτώνεται ολόκληρη η σελίδα.

- **Google Charts** (<https://developers.google.com/chart/>). Έτοιμα γραφήματα που μπορούν να ενσωματωθούν στις ιστοσελίδες, για την οπτικοποίηση των δεδομένων. Αντίστοιχα με τις διεπαφές και εδώ υπάρχουν αρκετά τρόποι οπτικοποίησης δεδομένων στο διαδίκτυο. Μετά την εξέταση αρκετών από αυτών καταλήξαμε στα γραφήματα που παρέχει η Google. Η επιλογή αυτή έγινε για λόγους αισθητικής καθώς επίσης και για τον λόγο ότι η Google παρείχε όλα τα γραφήματα που επιθυμούσαμε να χρησιμοποιήσουμε εξαρχής .
- **XML** (<http://www.w3.org/XML/>). Αποτελεί μια γλώσσα σήμανσης που ορίζει ένα σύνολο κανόνων για την κωδικοποίηση των δεδομένων. Σε XML είναι γραμμένη η απάντηση του server στον client.
- **JSON** (<http://www.json.org/>). Παρόμοια με XML αποτελεί μια γλώσσα σήμανσης με ένα διαφορετικό όμως σύνολο κανόνων για την κωδικοποίηση των δεδομένων. Σε JSON είναι γραμμένη η απάντηση του API της Διαύγειας στον server.
- **MY-SQL** (<http://www.sql.org/>). Βρίσκει εφαρμογή στην επικοινωνία του server με την βάση δεδομένων.

5.4.2 Πραγμάτωση

Κατά την διάρκεια δημιουργίας της εφαρμογής εντοπίστηκαν διάφορα θέματα υλοποίησης που έπρεπε να εξεταστούν. Τα βασικότερα ήταν:

- *Ποιών κυβερνητικών φορέων οι δαπάνες θα απεικονιστούν. Όλων;* Στο πρόγραμμα Διαύγεια είναι εγγεγραμμένοι 3816 ξεχωριστοί φορείς. Όπως είναι κατανοητό η απεικόνιση όλων των δαπανών από όλους τους φορείς θα προκαλούσε σύγχυση στο χρήστη. Έγινε η επιλογή να παρουσιάζονται μόνο οι δαπάνες των μεγαλύτερων δήμων, του Υπουργείου Τουρισμού ,του ΕΟΤ.

Επίσης *τι άλλα στοιχεία θα μπορούσαν να απεικονιστούν ;* Υπάρχει τεράστιος όγκος πιθανών δεδομένων προς απεικόνιση , και όπως είναι κατανοητό δεν μπορούν να απεικονιστούν όλα.

Γιαυτό επιλέξαμε στοιχεία τα οποία μπορεί εύκολα να τα κατανοήσει ο χρήστης και είναι ζητήματα σημαντικά για την κοινωνία μας. Οι πηγές, από τις οποίες συλλέχθηκαν τα δεδομένα είναι το ΣΕΤΕ, Η Πυροσβεστική και η Αστυνομία.

- *Ποία θα είναι η βασική δομή του ιστοχώρου;* Ακολουθώντας τις καινούργιες διαδικτυακές τάσεις επιλέξαμε να εστιάσουμε στην δημιουργία περισσότερο μιας διαδικτυακής εφαρμογής. Για να επιτύχουμε αυτόν τον σκοπό αποφασίσαμε η εφαρμογή να αποτελείται από 5 βασικές σελίδες. Την κεντρική σελίδα που υπάρχει μία εισαγωγή ,την δεύτερη σελίδα που απεικονίζοντε τα δεδομένα που συλλέχθηκαν από το Διαύγεια ,την τρίτη σελίδα που απεικονίζοντε στοιχεία από το ΣΕΤΕ ,την τέταρτη σελίδα που απεικονίζοντε στοιχεία από την Πυροσβεστική και την Πέμπτη σελίδα που απεικονίζοντε στοιχεία από την Αστυνομία . Έτσι ο χρήστης έχει 5 βασικά ξεκάθαρα στιγμιότυπα και δεν αναλώνεται στην περιήγηση.

- *Τι οπτικοποιήσεις θα χρησιμοποιηθούν ;* Όπως είναι κατανοητό η όλων των οπτικοποιήσεων θα προκαλούσε σύγχυση στο χρήστη.Έτσι αποφασίσαμε να χρησιμοποιήσουμε οπτικοποιήσεις που μπορεί να τις καταλάβει και να τις κατανοήσει ο καθένας και να μπορεί να αντλίσει εύκολα σημαντική πληροφορία.


5.4.3 Περιπτώσεις Χρήσης

Η εφαρμογή μας, η οποία φιλοξενείται στην διεύθυνση www.greekopendata.host22.com/, αποτελείται από 5 σελίδες.

Παρακάτω θα αναλύσουμε κάθε σελίδα ξεχωριστά

✚ Στην πρώτη σελίδα (Αρχική) περιέχεται ένα κείμενο σχετικό με τη σημασία των ανοιχτών δεδομένων, αλλά και με την εφαρμογή. Ο στόχος της Αρχική μας σελίδας είναι να καλωσορίσει τον επισκέπτη και να τον βοηθήσει να κατανοήσει την σημασία για την δημιουργία της εφαρμογής.

Ανοιχτά Δημόσια Δεδομένα



Αρχική | Διαύγεια | ΣΕΤΕ | Πυροσβεστική | Αστυνομία

Καλώς Ηρθατε

Τα *Ανοιχτά Δεδομένα* και ειδικότερα τα *Ανοιχτά Κυβερνητικά Δεδομένα*, είναι μια τεράστια πηγή δεδομένων που δεν έχει ακόμα αξιοποιηθεί. Πολλά μεμονωμένα άτομα, αλλά και οργανισμοί, συλλέγουν ένα ευρύ φάσμα διαφορετικών τύπων δεδομένων, προκειμένου να εκπληρώσουν τις εργασίες τους. Τα κυβερνητικά δεδομένα είναι ιδιαίτερα σημαντικά από αυτήν την άποψη, τόσο λόγω της ποσότητας και της καθολικότητάς τους, όσο και λόγω του ότι αυτά τα δεδομένα είναι, από νομικής άποψης, δημόσια δεδομένα και συνεπώς θα μπορούσαν να είναι ανοιχτά και διαθέσιμα σε άλλους να τα χρησιμοποιήσουν.

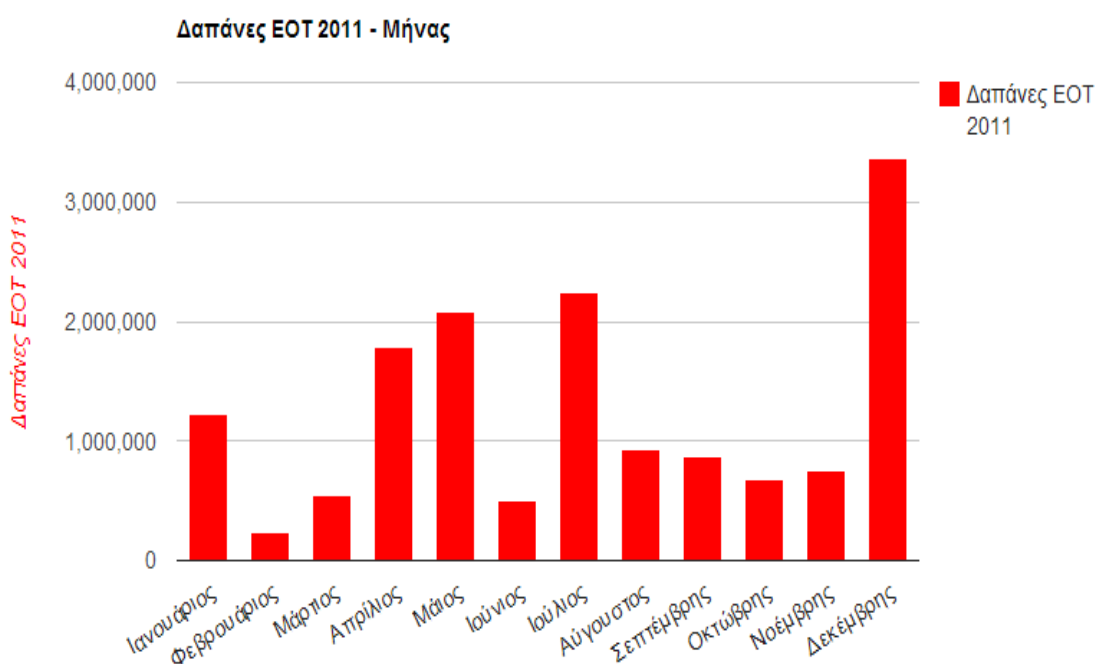
Ο Νόμος 3979/2011 (Νόμος για την Ηλεκτρονική Διακυβέρνηση) προβλέπει την ελεύθερη διάθεση των δεδομένων χωρίς ουσιαστικούς περιορισμούς, με τρόπο δηλαδή που επιτρέπει την περαιτέρω χρήση και επεξεργασία τους από τον καθένα. Όμως αυτό που κρίνει τελικά την ουσία του δικαιώματος δεν είναι μόνο το νομικό πλαίσιο, αλλά επίσης, οι οργανωτικές και τεχνολογικές επιλογές που το υλοποιούν. Στο τομέα αυτό απαιτούνται περαιτέρω βήματα και συντονισμός καθώς η Ελλάδα υστερεί σε σχέση με άλλες χώρες.

Υπάρχουν πολλοί τομείς που μπορούμε να περιμένουμε πως τα *Ανοιχτά Δεδομένα* μπορούν να έχουν αξία και υπάρχουν ήδη παραδείγματα που αναδεικνύουν πως αυτά έχουν χρησιμοποιηθεί. Υπάρχουν, επίσης, πολλές ομάδες ανθρώπων και οργανισμών που μπορούν να επωφεληθούν από τη διαθεσιμότητα των ανοιχτών δεδομένων, συμπεριλαμβανομένων των ίδιων των κυβερνήσεων. Την ίδια στιγμή, είναι αδύνατον να προβλέψουμε με ακρίβεια πώς και πού μπορούν να προσδώσουν αξία μελλοντικά. Σε αυτήν την εφαρμογή γίνεται μια προσπάθεια για την αξιοποίηση Δημόσιων Ανοιχτών Δεδομένων τόσο με στατικό, όσο και με δυναμικό τρόπο, έτσι ώστε να αυξηθεί η πληροφοριακή τους αξία. Οι πηγές, από τις οποίες συλλέχθηκαν τα δεδομένα είναι το Διαύγεια (μέσω των APIs του Διαύγεια), το ΣΕΤΕ, Η Πυροσβεστική και η Αστυνομία.

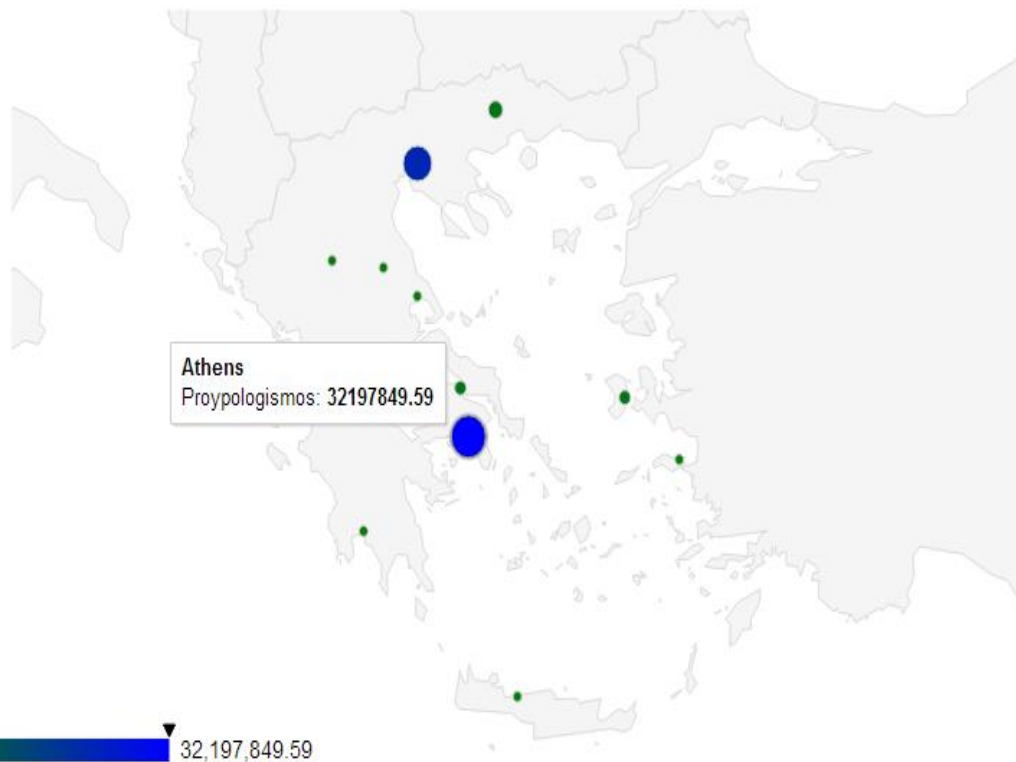
✚ Στην δεύτερη σελίδα(Διαύγεια) η εφαρμογή μας επιτρέπει την παρακολούθηση των αποφάσεων φορέων του Ελληνικού Δημοσίου.

Για την καλύτερη διαξαγωγή αποτελεσμάτων έχουν χρησιμοποιηθεί συγκεκριμένες οπτικοποιήσεις. Η υλοποίηση της βασίστηκε στο opendata API του Διαύγεια και στα google charts. Για να μάθετε περισσότερα για το API του Διαύγεια επισκεφτείτε opendata.diavgeia.gov.gr.

Όλες οι οπτικοποιήσεις που χρησιμοποιούνται ενημερώνονται δυναμικά κατά τακτά χρονικά διαστήματα, έτσι όπως τα έχουμε ορίσει. Είναι σημαντικό να αναφέρουμε ότι τα δεδομένα που παρουσιάζονται σε αυτό το κομμάτι της εφαρμογής μας μπορεί να διαφέρουν από την πραγματικότητα. Αυτό οφείλεται στην αναποτελεσματική συνεργασία του Διαύγεια με κάποιους από τους φορείς του Δημοσίου. Παρακάτω μπορείτε να δείτε κάποιες από τις οπτικοποιήσεις που χρησιμοποιήσαμε :



Προϋπολογισμός μεγαλύτερων Δήμων 2012 ▾



Δαπάνες-Προϋπολογισμός δήμων 2012 ▾

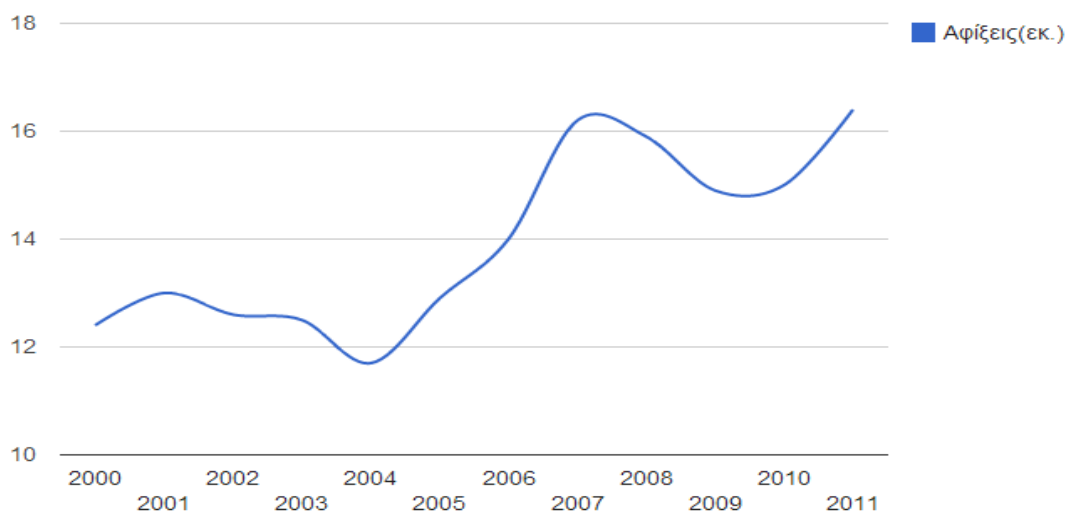
Πόλη	Δαπάνες	Προϋπολογισμός
Athens	4171	32197849.59
Thessaloniki	4810650.95	22525145.34
Larissa	112602.44	394528.84
Drama	599883.54	2919225.46
Kavala	1070685.09	2618893.01
Volos	1996.82	706310.36
Kalampaka	409536.58	0
Patra	239721.55	2434582.26
Kalamata	1034804.59	318108.93
Rethimno	1066.56	516291.13
Samos	607247.68	0
Chios	1394239.29	1951438.32

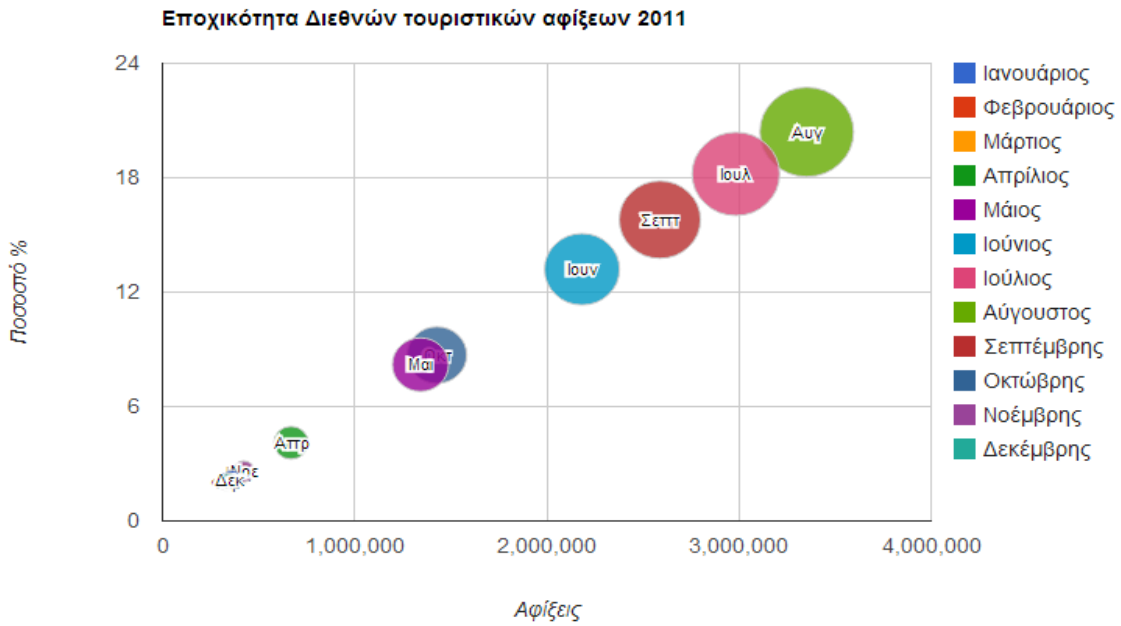
✚ Στην κατηγορία "Σετε" περιέχονται οι κατηγορίες **αφίξεις τουριστών** και **έσοδα τουρισμού** καθεμιά από τις οποίες μας οδηγούν σε ξεχωριστές σελίδες με αντίστοιχες οπτικοποιήσεις. Τα δεδομένα που χρησιμοποιήθηκαν έχουν συλλεχθεί από τον Σύνδεσμο Ελληνικών Τουριστικών Επιχειρήσεων (<http://www.sete.gr/>) αλλά και από Ελληνική Στατιστική Αρχή (<http://www.statistics.gr>).

- Για «Αφίξεις Τουριστών» :

Στην κατηγορία αυτή περιέχονται οπτικοποιήσεις σχετικές με τις αφίξεις των τουριστών στην Ελλάδα. Στην κάθε κατηγορία δίνεται επιλογή για συγκεκριμένη οπτικοποίηση.

Παρακάτω μπορείτε να δείτε κάποιες από τις οπτικοποιήσεις που χρησιμοποιήσαμε :

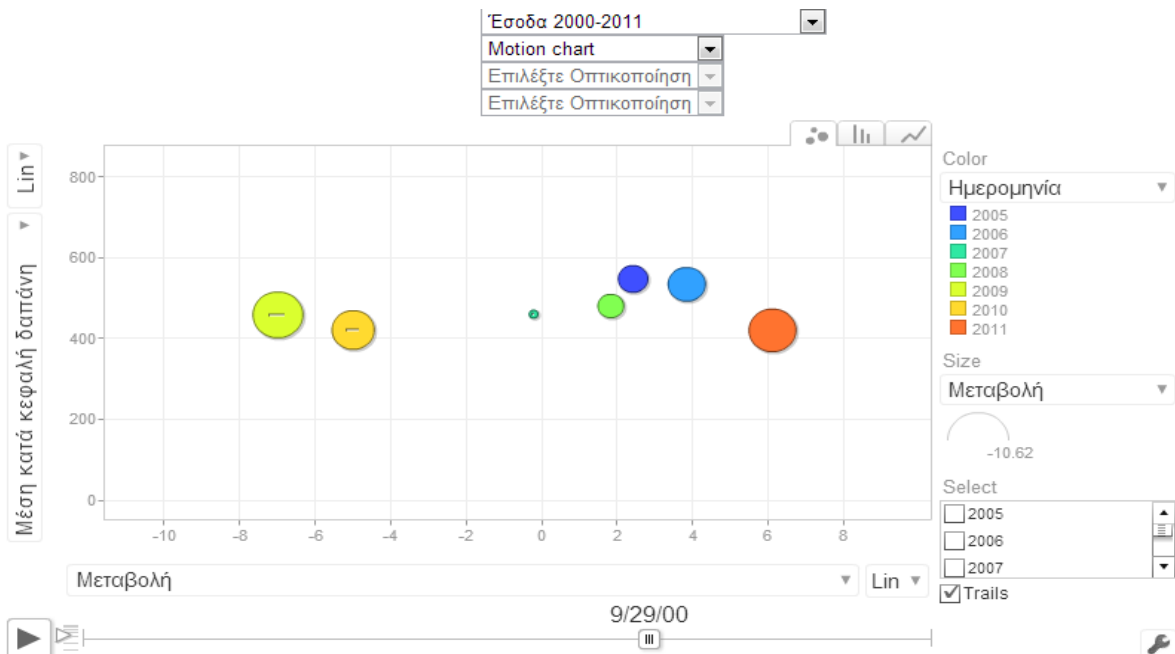




- Για « Έσοδα Τουρισμού » :

Στην κατηγορία "Έσοδα Τουρισμού" περιέχονται οπτικοποιήσεις σχετικές με της Ελλάδας από τον Τουρισμό. Στην κάθε κατηγορία δίνεται επιλογή για συγκεκριμένη οπτικοποίηση.

Παρακάτω μπορείτε να δείτε κάποιες από τις οπτικοποιήσεις που χρησιμοποιήσαμε :



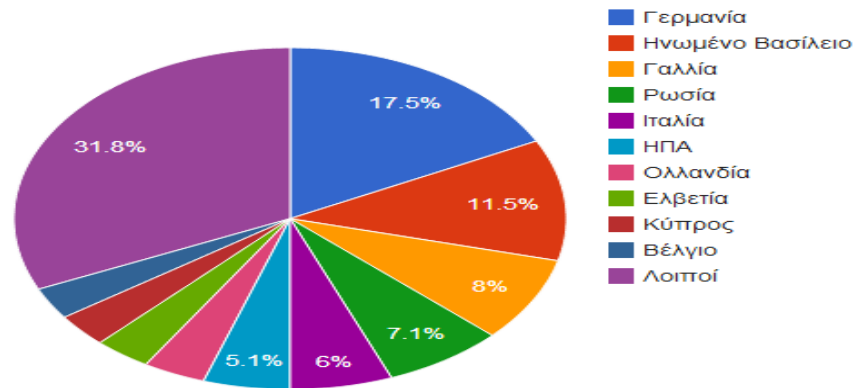
Έσοδα ανά χώρα προέλευσης 2011

Motion chart

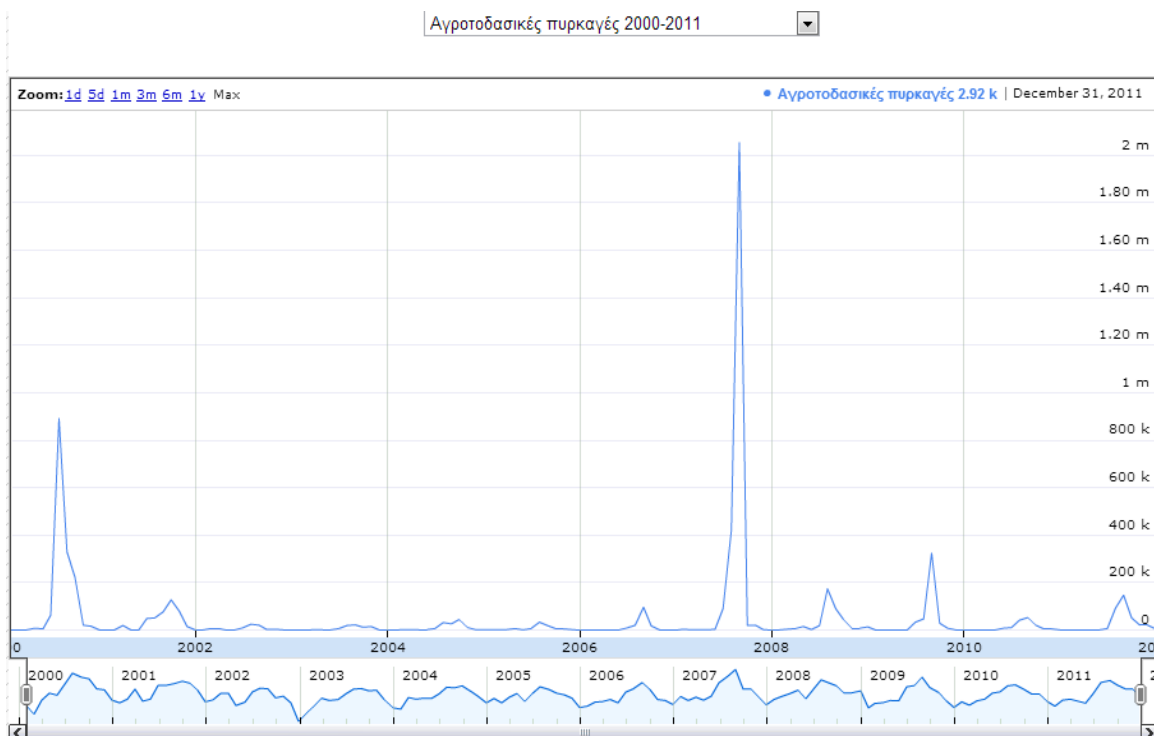
Pie chart

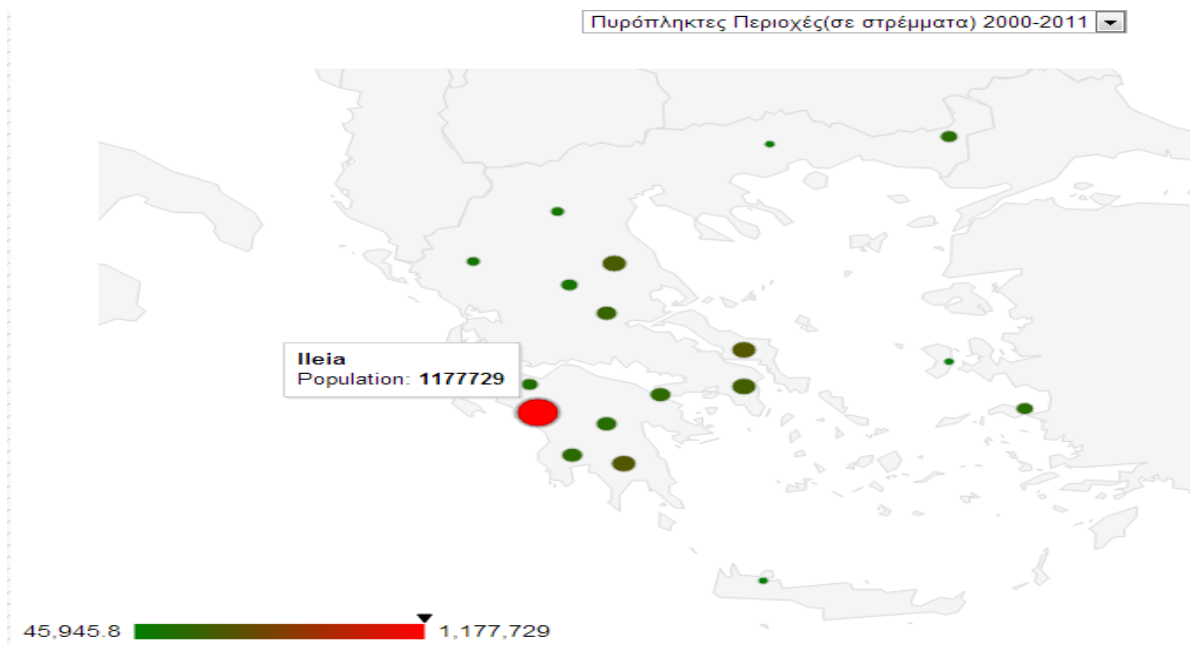
Επιλέξτε Οπτικοποίηση

Έσοδα ανά χώρα προέλευσης 2011

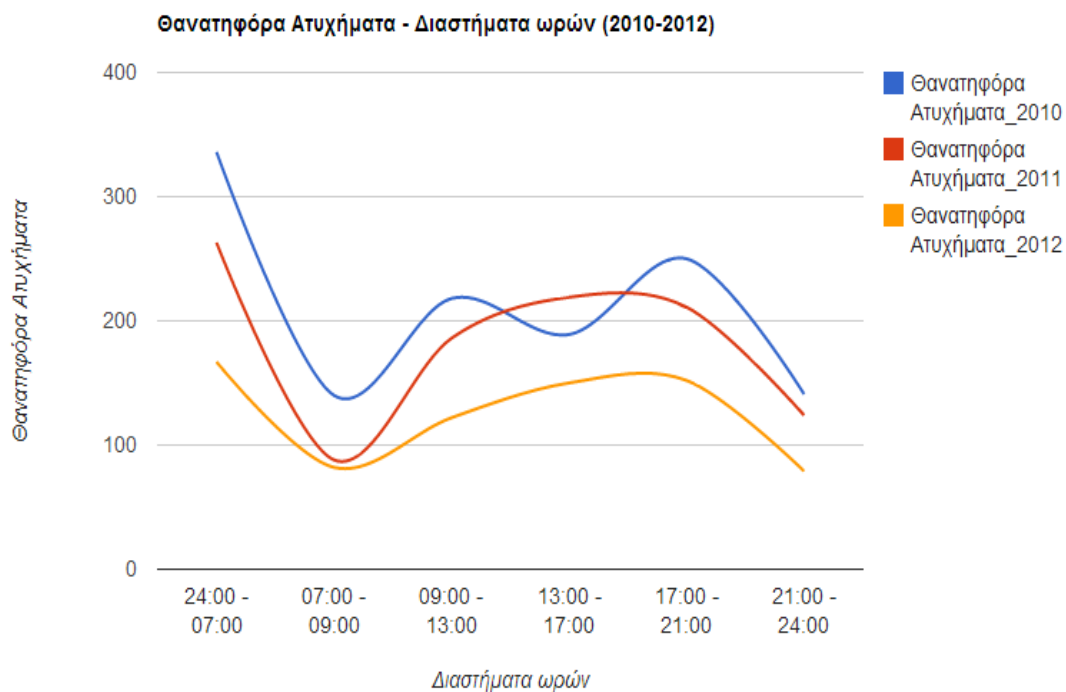


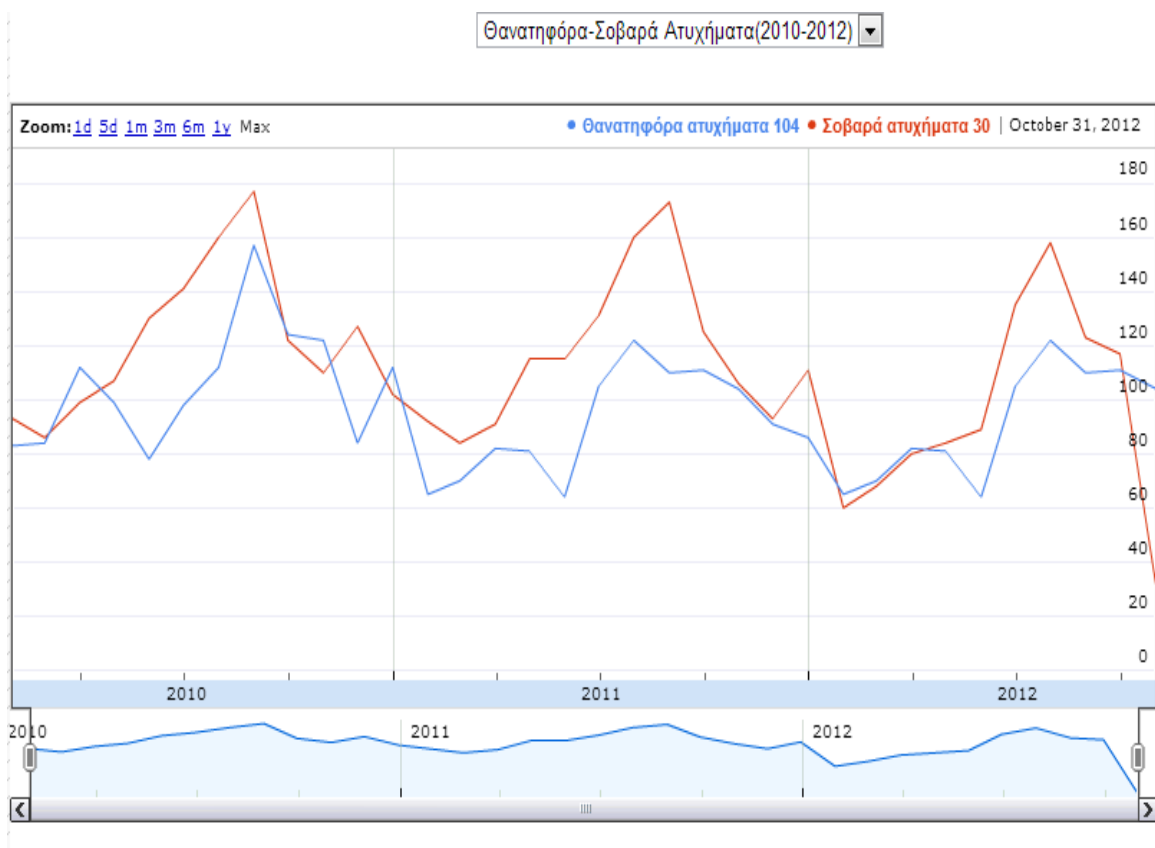
✚ Στην κατηγορία "Πυροσβεστική" περιέχονται οπτικοποιήσεις σχετικές με τα αστικά συμβάντα και τις πυρκαγές που συνέβησαν στην Ελλάδα στο χρονικό διάστημα 2000-2011. Τα δεδομένα που χρησιμοποιήθηκαν έχουν συλλεχθεί από το Πυροσβεστικό Σώμα Ελλάδος (<http://www.fireservice.gr>) και κατηγοριοποιούνται ανάλογα με το είδος, την χρονολογία και το μέρος του συμβάντος. Παρακάτω μπορείτε να δείτε κάποιες από τις οπτικοποιήσεις που χρησιμοποιήσαμε :





✚ Στην κατηγορία "**Αστυνομία**" περιέχονται οπτικοποιήσεις σχετικές με τις τα ατυχήματα που συνέβησαν στην Ελλάδα στο χρονικό διάστημα 2000-2012. Τα δεδομένα που χρησιμοποιήθηκαν έχουν συλλεχθεί από την Ελληνική Αστυνομία (<http://www.astynomia.gr/>) και κατηγοριοποιούνται ανάλογα με την χρονολογία το είδος και την σοβαρότητά τους. Παρακάτω μπορείτε να δείτε κάποιες από τις οπτικοποιήσεις που χρησιμοποιήσαμε :





✚ Στην κατηγορία "ΕΛ.ΣΤΑΤ." περιέχονται οπτικοποιήσεις σχετικές με την ανεργία και με κοινωνικά στοιχεία, όπως ο πληθυσμός των μόνιμων μεταναστών στην χώρα μας σύμφωνα με την απογραφή της ΕΛ.ΣΤΑΤ. το 2011. Τα δεδομένα που χρησιμοποιήθηκαν έχουν συλλεχθεί από το Ελληνική Στατιστική Υπηρεσία της Ελλάδος (<http://www.statistics.gr/>). Παρακάτω μπορείτε να δείτε κάποιες από τις οπτικοποιήσεις που χρησιμοποιήσαμε:

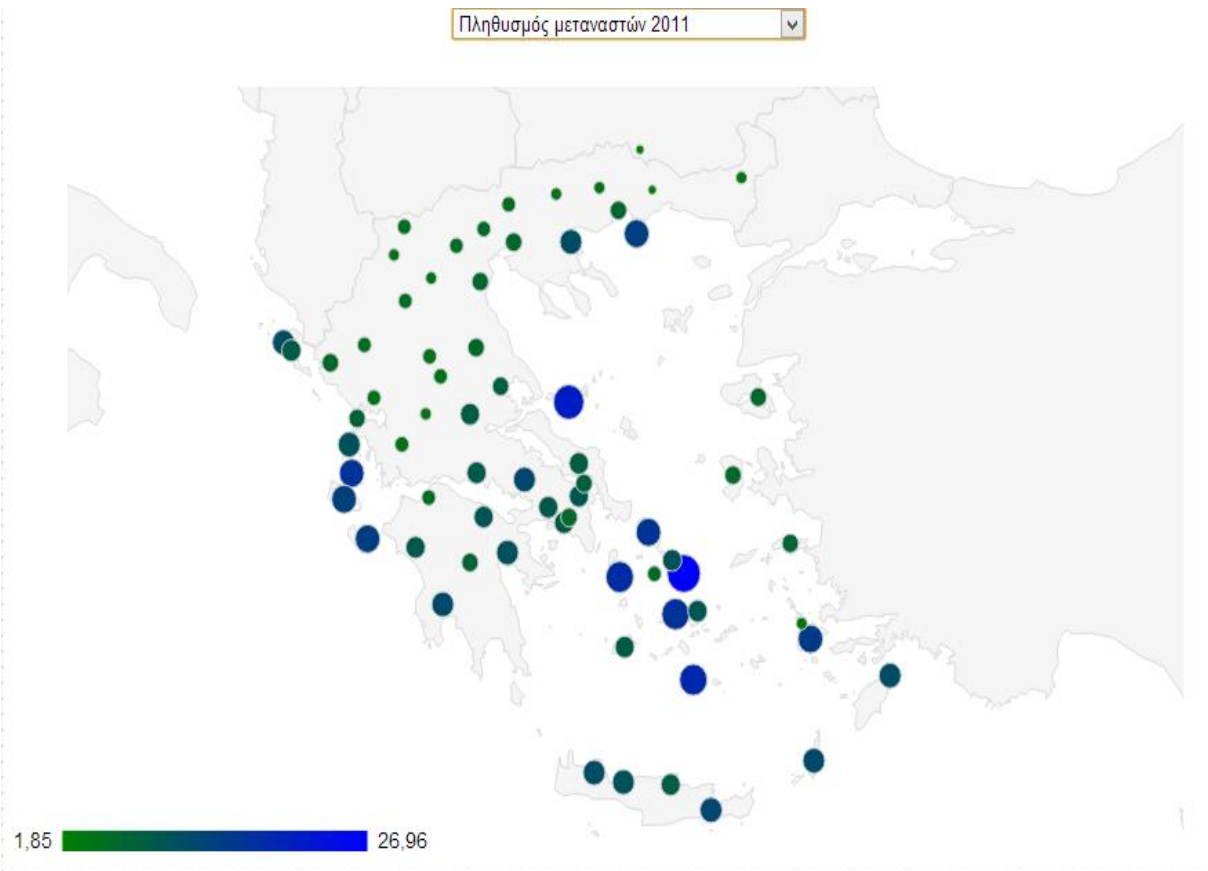
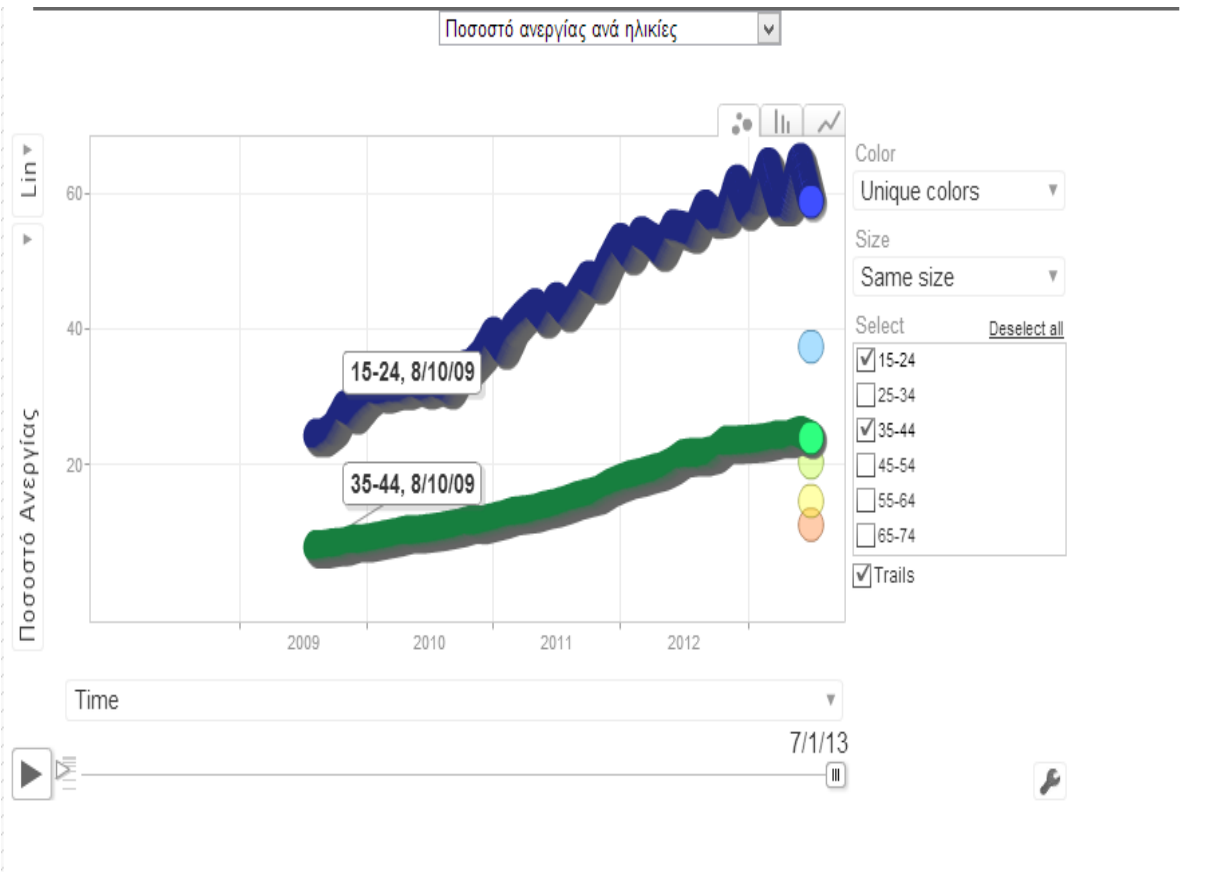
***Ανοιχτά Δημόσια
Δεδομένα***

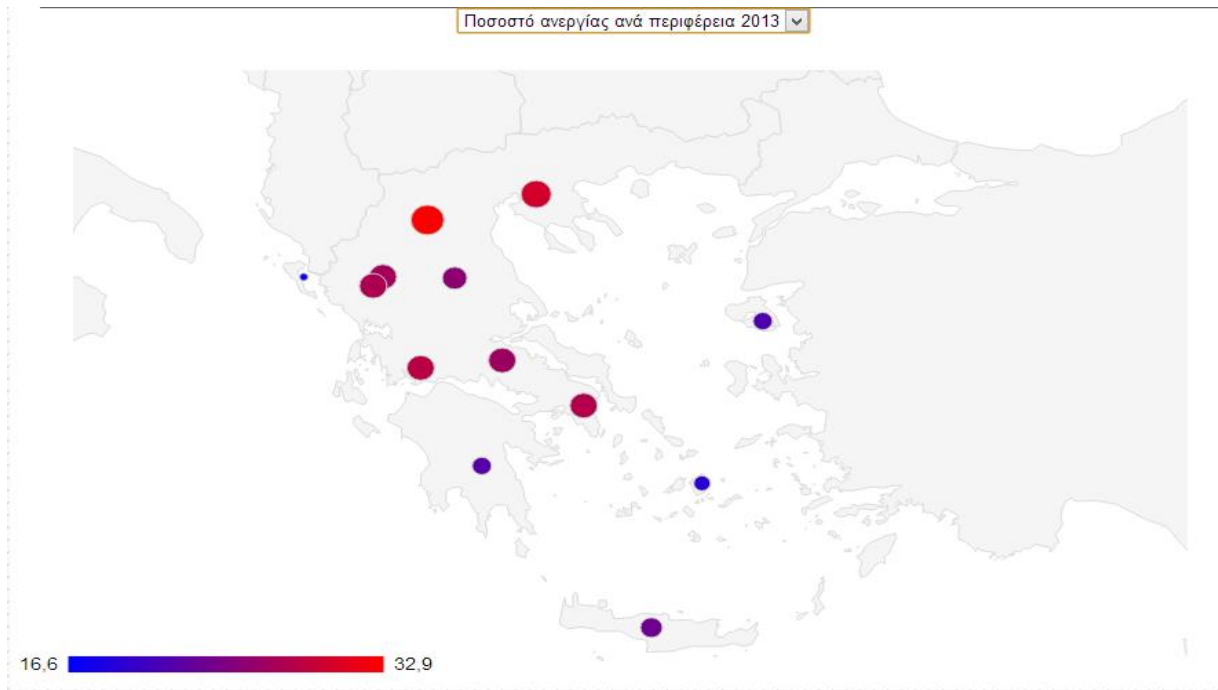
[Αρχική](#) | [Διαύγεια](#) | [ΣΕΤΕ](#) | [Πυροσβεστική](#) | [Αστυνομία](#) | [ΕΛ.ΣΤΑΤ](#) | [EUROSTAT](#)

ΕΛ.ΣΤΑΤ.

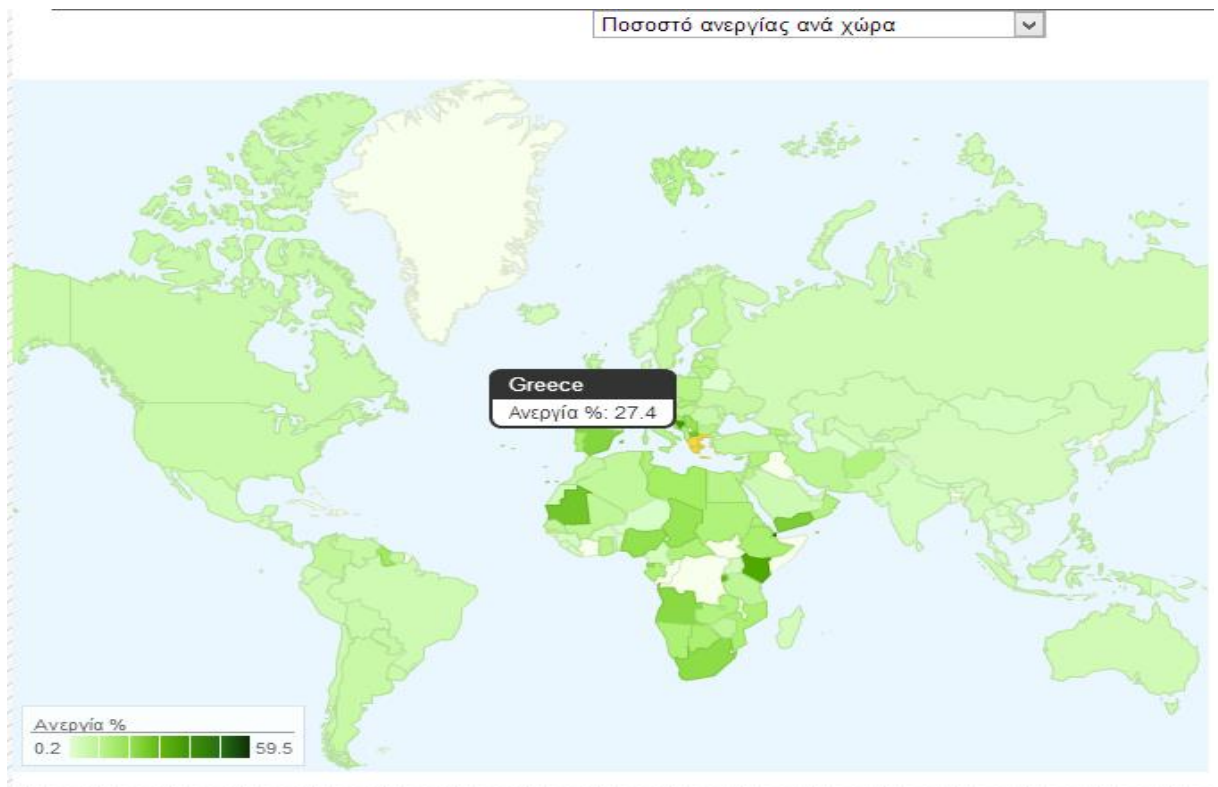
Στην κατηγορία "ΕΛ.ΣΤΑΤ." περιέχονται οπτικοποιήσεις σχετικές με την ανεργία και με κοινωνικά στοιχεία, όπως ο πληθυσμός των μόνιμων μεταναστών στην χώρα μας σύμφωνα με την απογραφή της ΕΛ.ΣΤΑΤ. το 2011. Τα δεδομένα που χρησιμοποιήθηκαν έχουν συλλεχθεί από το Ελληνική Στατιστική Υπηρεσία της Ελλάδος (<http://www.statistics.gr/>). Μπορείτε παρακάτω να επιλέξετε την οπτικοποίηση, που επιθυμείτε να δείτε :

▾



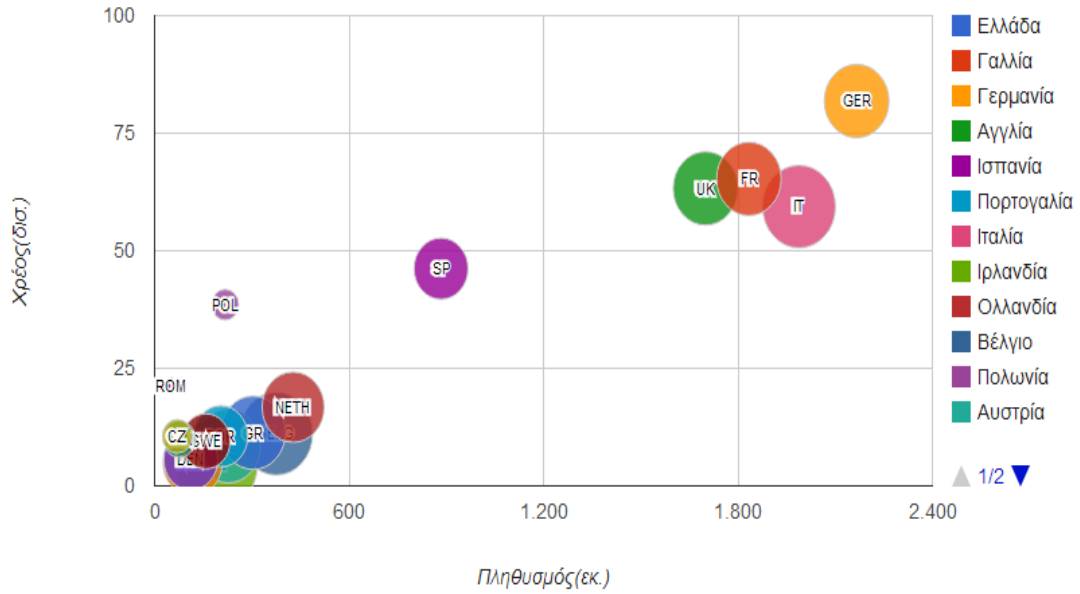


✚ Στην κατηγορία "EUROSTAT." περιέχονται οπτικοποιήσεις σχετικές με την παγκόσμια ανεργία και με οικονομικά στοιχεία, όπως ο κατώτατος μισθός ανά χώρα ή το δημόσιο χρέος κάθε χώρας. Τα δεδομένα που χρησιμοποιήθηκαν έχουν συλλεχθεί από το EUROSTAT (<http://err.eurostat.ec.europa.eu/>). Παρακάτω μπορείτε να δείτε κάποιες από τις οπτικοποιήσεις που χρησιμοποιήσαμε:

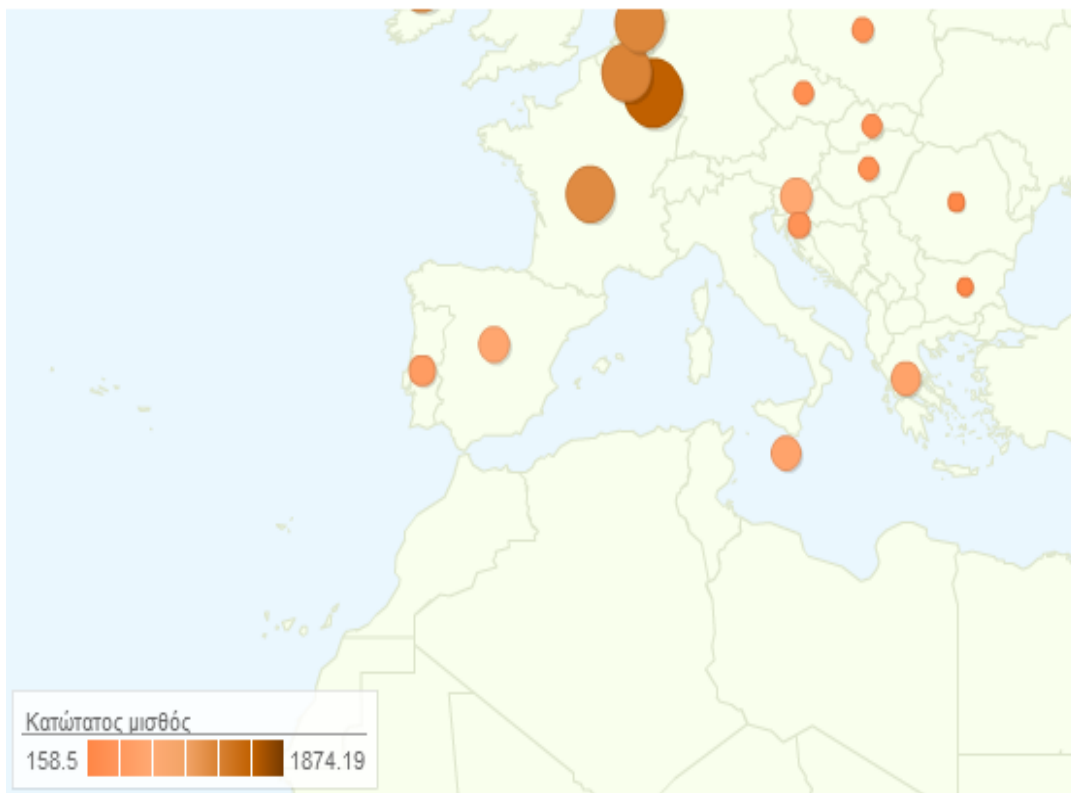


Το χρέος των χωρών της Ευρώπης 2013 ▾

Στατιστικά στοιχεία για το χρέος των χωρών της Ευρώπης. Το μέγεθος των bubbles απεικονίζει το χρέος της κάθε χώρας σε σχέση με τον πληθυσμό της



Κατώτατος μισθός στην Ευρώπη 2013 ▾



ΚΕΦΑΛΑΙΟ 6 – Συμπεράσματα

6.1 Σύνοψη - Συμπεράσματα

Στα πλαίσια αυτής της εργασίας ασχοληθήκαμε με τα ανοιχτά κυβερνητικά δεδομένα που προσφέρει η ελληνική κυβέρνηση μέσω του προγράμματος Διαύγεια, του ΣΕΤΕ, της Πυροσβεστικής, της Αστυνομίας, της Ελληνικής Στατιστικής Υπηρεσίας και του Eurostat. Ξεκινήσαμε από το γενικό, τα ανοιχτά δεδομένα, και καταλήξαμε στο ειδικό, τα κυβερνητικά ανοιχτά δεδομένα της Ελλάδας. Βασικό αντικείμενο της μελέτης μας ήταν ο καταλληλότερος τρόπος οπτικοποίησης των δεδομένων που είχαμε. Αυτό προϋπόθετε την εξέταση των δεδομένων που μας παρείχαν και στον εντοπισμό των πληροφοριών που μπορούσαμε να εξάγουμε. Παρουσιάστηκαν, συνοπτικά, αρκετά εργαλεία οπτικοποίησης αλλά στο τέλος η καταλληλότερη επιλογή για εμάς ήταν τα Google Charts.

Επίσης προβλήθηκαν όλες οι διαθέσιμες παρόμοιες εργασίες. Μελετήθηκαν τα χαρακτηριστικά τους και εντοπίστηκαν τα μειονεκτήματά τους. Προσπαθήσαμε να τα βελτιώσουμε τόσο όσο και να συμπληρώσουμε όποια κενά διακρίναμε ότι υπήρχαν. Ο βασικός μας στόχος για την προώθηση της συμμετοχικής δημοκρατίας πιστεύουμε ότι είναι επιτυχημένος.

Ένα από τα σημαντικότερα συμπεράσματά μας, ύστερα από την υλοποίηση της εργασίας αφορά την διαφάνεια και την σημαντικότητα των Ανοιχτών Κυβερνητικών Δεδομένων. Όπως αναφέραμε και σε προηγούμενο κεφάλαιο, έχουν πραγματοποιηθεί σημαντικά βήματα με στόχο την ενημέρωση και την διαφάνεια των κυβερνητικών αποφάσεων. Ίσως το σημαντικότερο βήμα σε αυτή την προσπάθεια είναι το πρόγραμμα Διαύγεια. Παρόλα αυτά κάποια από τα προβλήματα που παρουσιάζονται σε αυτόν τον μηχανισμό, καθιστούν αδύνατη την σωστή ανάλυση των δεδομένων που προκύπτουν. Αυτό έχει σαν αποτέλεσμα να μην επιτυγχάνεται ο αρχικός στόχος αυτής της προσπάθειας. Συμπερασματικά, είναι αναγκαίο κατά τη γνώμη μας να δοθεί μεγαλύτερη έμφαση σε προβλήματα όπως αυτά που αναφέραμε σε προηγούμενο κεφάλαιο και αφορούν το πρόγραμμα του Διαύγεια.

6.2 Πηγές - Βιβλιογραφία

[1] Παγκόσμιος Ιστός

http://users.forthnet.gr/ath/skonstan/site_1/articles/history_files/Web.html

[2] Linked Data: Principles and State of the Art

<http://www.w3.org/2008/Talks/WWW2008-W3CTrack-LOD.pdf>

[3] Future of the Web, Tim Berners-Lee

<http://net.educause.edu/ir/library/pdf/EPO0719.pdf>

[4] Technical Report on Linked Data Applications - The Genesis and the Challenges of Using Linked Data on the Web

http://linkeddata.deri.ie/sites/linkeddata.deri.ie/files/lod-app-tr-2009-07-26_0.pdf

[5] Exploiting Linked Data for Building Web Applications, Michael Hausenblas

<http://sw-app.org/pub/exploit-lod-webapps-IEEEIC-preprint.pdf>

[6] Linked Data – The story so far

<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

[7] How to publish Link Data on the Web (tutorial)

<http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>

[8] Cool URIs for the Semantic Web

<http://www.dfki.uni-kl.de/~sauermann/2006/11/cooluris/>

[9] Tony Segaran, Jeff Hammerbacher, Beautiful Data, O'Reilly, August 2009 (page 335)

[10] Interlinking Open Data on the Web

<http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkingOpenData.pdf>

[11] A Practical Introduction by Michael Hausenblas

http://www.deri.ie/fileadmin/documents/teaching/tutorials/LinkedData_MichaelHausenblas_2009-03-05.pdf

[12] Putting Government Data Online, Tim Berners-Lee

<http://www.w3.org/DesignIssues/GovData.html>

[13] Publishing Open Government Data - W3C Working Draft 8 September 2009

<http://www.w3.org/TR/2009/WD-gov-data-20090908/>

[14] Open Government Data

<http://www.w3.org/TR/2009/NOTE-egov-improving-20090512/#OGD>

[15] Δημόσια Δεδομένα, Δικά μας Δεδομένα

<http://www.publicdata.gr/>

[16] Open Government Data Standards and Setting Expectations

http://razor.occams.info/pubdocs/2009-02-28_TCamp_Data_Standards.pdf

[17] Linking Open Data: An Emerging Practice Area for the Semantic Web

<http://phaneron.rickmurphy.org/?p=34>

[18] Government Linked Data: A Tipping Point for the Semantic Web

http://eprints.ecs.soton.ac.uk/18311/1/GLD-A_Tipping_point_for_the_SW_Shanghai_Keynote.pdf

[19] Ordnance Survey, UK

<http://data.ordnancesurvey.co.uk>

[20] Τσιαβός Π., Ζητήματα Ανοιχτής Διάθεσης και Διαμοιρασμού Δεδομένων στο Ν. 3882/2010, 2011, άρθρο

http://www.inspire.okxe.gr/assets/docs/attachments/presentations/inep/presentation_Prodromos_Tsiavos.pdf

[21] Δεληβοριάς Χ., *Life is a State of Mind, 2010*, ηλεκτρονικό άρθρο

<http://delivorias.me/2010/01/31/ανοιχτά-δεδομένα-linked-data-rdfa/>

[22] Ακριώτης Ιωάννης, “Εκτίμηση του Προσδοκώμενου Κέρδους από το Μετασχηματισμό και την Ολοκλήρωση Δημοσίων Υπηρεσιών”, ΔΠΜΣ “Τεχνοοικονομικά Συστήματα”, Εθνικό Μετσόβιο Πολυτεχνείο, Ιούνιος 2010

[23] Αλεξόπουλος Χ., Διαλειτουργικότητα - Η δύναμη που μετασχηματίζει οργανισμούς, δεδομένα και πληροφοριακά Συστήματα, Διπλωματική εργασία, Μεταπτυχιακό πρόγραμμα σπουδών – Διοίκηση πληροφοριακών συστημάτων, Τμήμα Μηχανικών και Επικοινωνιακών συστημάτων, Πανεπιστήμιο Αιγαίου, 2010

<https://pithos.grnet.gr/pithos/rest/irematz@ntua.gr/files/DIPLOMATIKES/xalexopoulos.pdf>

[24] Πλαίσιο Διαλειτουργικότητας & Υπηρεσιών Ηλεκτρονικών Συναλλαγών

Έκδοση 4.0, Μάρτιος 2012

<http://www.yap.gov.gr/images/stories/Plaisio/PDYHS-v4.0.pdf>

[25] Σπηλιωτοπούλου Λ., Επισκόπηση των δυνατοτήτων των ηλεκτρονικών υπηρεσιών που διαθέτουν οι διαδικτυακοί τόποι των Υπουργείων της Ελληνικής Κυβέρνησης, Μεταπτυχιακό πρόγραμμα σπουδών – Διοίκηση πληροφοριακών συστημάτων, Τμήμα Μηχανικών και Επικοινωνιακών συστημάτων, Πανεπιστήμιο Αιγαίου, 2011.

[26] Παπαδόπουλος Θωδωρής. (2011) . Ανοιχτά Δημόσια Δεδομένα , Εθνική Σχολή Δημόσιας Διοίκησης , διαθέσιμη από <http://thodoris.net/files/OpenPublicData.pdf>