

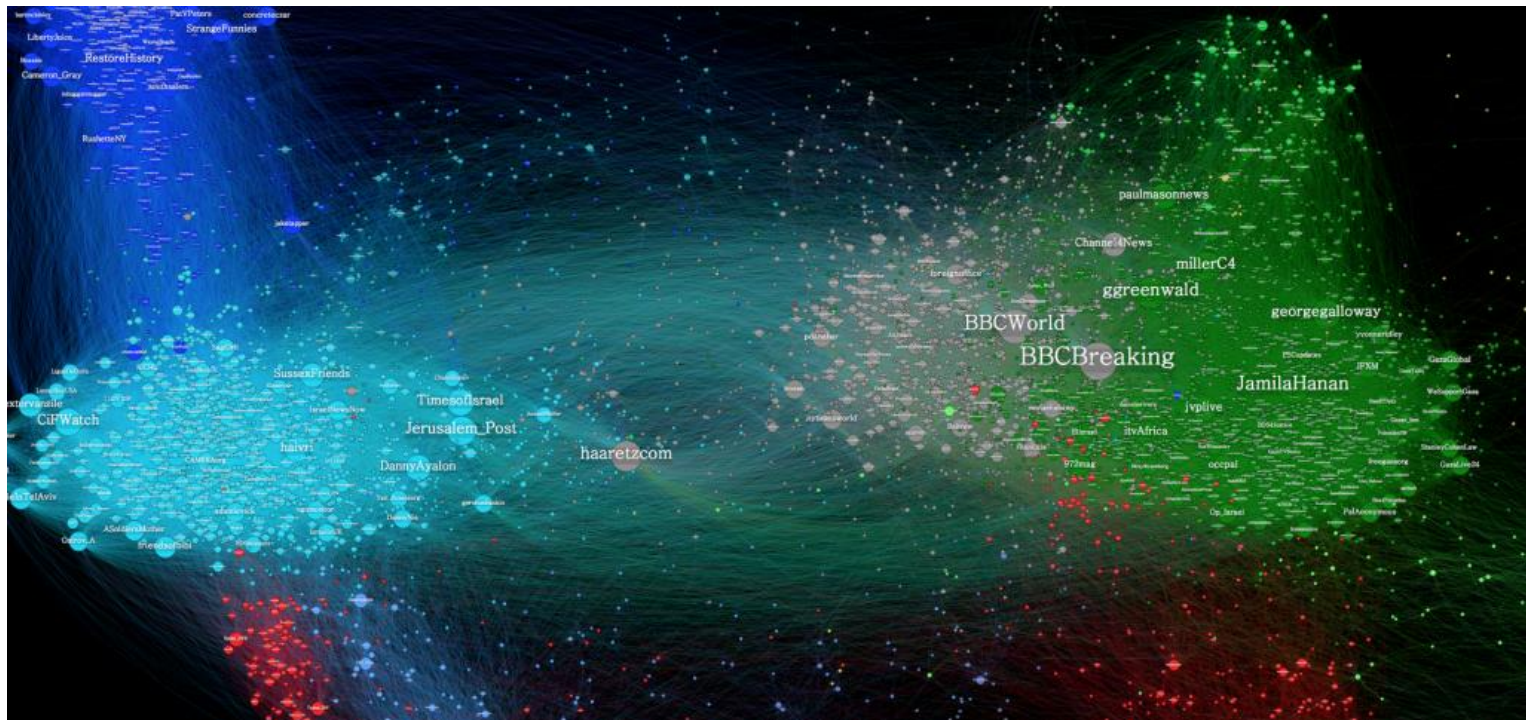
Τι είναι τα μεγάλα δεδομένα

- Η συγκέντρωση μεγάλου όγκου δεδομένων από ψηφιακές εφαρμογές, υπηρεσίες και πάσης φύσης συσκευές που είναι συνδεδεμένες στο διαδίκτυο των πραγμάτων (IOT).
- με το συγκεκριμένο όρο αναφερόμαστε συνήθως σε όγκους δεδομένων που κυμαίνονται από μερικά terabytes έως δεκάδες ή και εκατοντάδες petabytes (1.024 terabytes) ή exabytes (1.024 petabytes) ή zetabytes(1.024 exabytes).

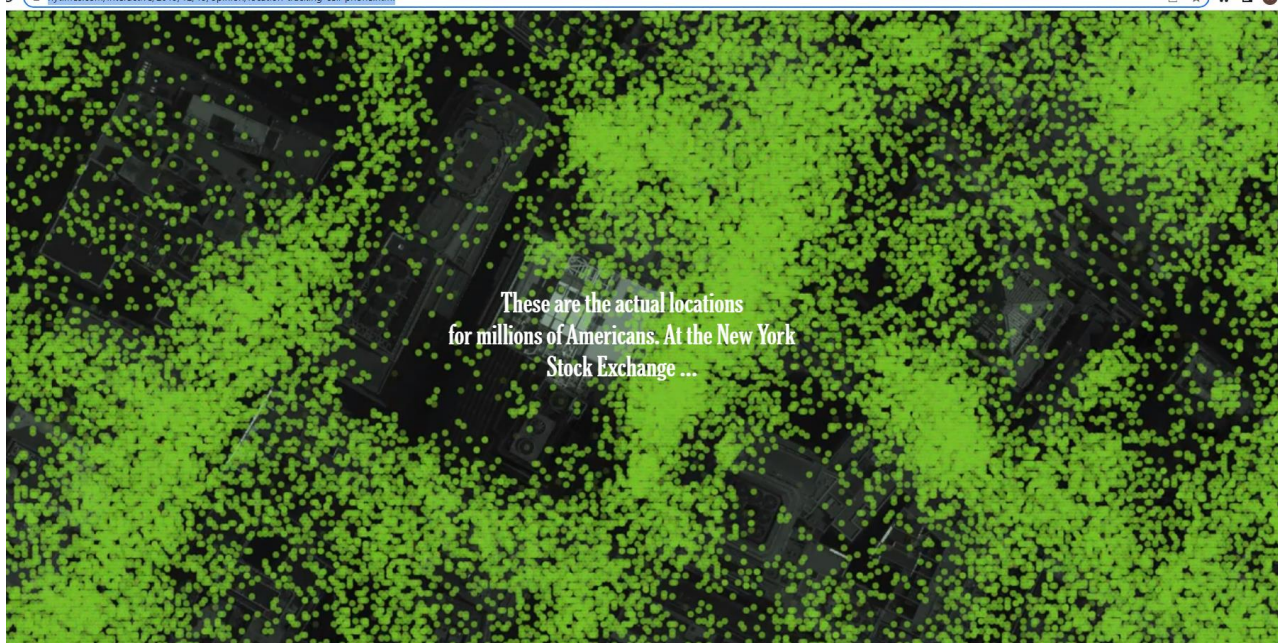
Τι είναι τα μεγάλα δεδομένα

- Επίσης αναφερόμαστε στις υποδομές και τεχνολογίες αποθήκευσης των δεδομένων αυτών καθώς και σε εφαρμογές οι οποίες μπορούν να κάνουν
 - Στατιστική ανάλυση στα δεδομένα αυτά
 - Οπτικοποιήσεις
 - Εξαγωγή γνώσης (με αλγόριθμους εξόρυξης γνώσης)

Twitter visualizations



Μεγάλα δεδομένα



- <https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html>

Από πού προέρχονται τα δεδομένα

- Πολλές τεχνολογικές καινοτομίες έχουν οδηγήσει σε δραματική αύξηση των δεδομένων και στη συλλογή δεδομένων
- Δεν είναι όμως μόνο οι οργανισμοί που παράγουν τεράστιους όγκους δεδομένων.
- Ακόμη και σε μικρότερη κλίμακα οργάνωσης, στο επίπεδο του ατόμου, η παραγωγή δεδομένων είναι πρωτόγνωρη (π.χ. αναζητήσεις στις μηχανές αναζήτησης, χρήση social media, online πλοήγηση και αγορές σε εμπορικά sites κ.α. χρήση υπηρεσιών κτλ)

Παραδείγματα μεγάλων δεδομένων

- δεδομένα από εταιρικές εφαρμογές (π.χ. CRM, ERP συστήματα)
- δεδομένα από κρατικές υπηρεσίες (π.χ. Taxisnet)
- δεδομένα από online εμπορικές εφαρμογές (π.χ. Amazon, eBay)
- δεδομένα από εκπαιδευτικές πλατφόρμες (π.χ. Coursera, eDX, KhanAcademy)
- δεδομένα από την χρήση μηχανών αναζήτησης Google
- Δεδομένα από την χρήση υπηρεσιών (π.χ. Google drive, YouTube)
- Δεδομένα από την χρήση των Social Media (Facebook, Twitter, Instagram)

Επίσης...

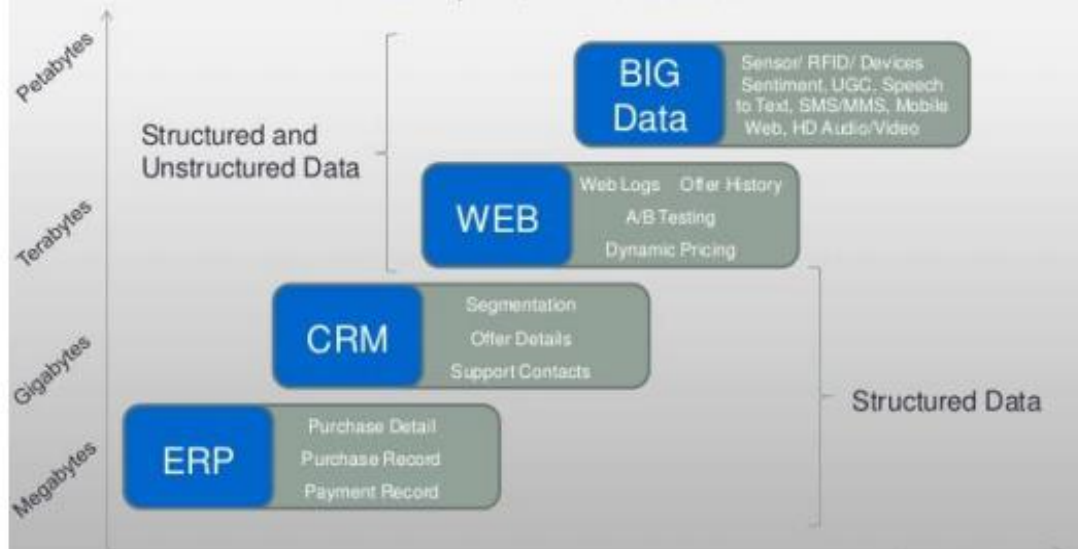
- Τομείς, όπως η ιατρική, η αστρονομία, η μετεωρολογία αλλά και η βιολογία χάρη στις νέες τεχνολογίες, τα νέα τηλεσκόπια, τους νέους και φτηνούς αισθητήρες και τα νέα μηχανήματα για την αποκωδικοποίηση DNA μπορούν και παράγουν τεράστιους όγκους δεδομένων.

Επίσης..

- Μεγάλα δεδομένα επίσης παράγει η
 - Χρήση κινητών τηλεφώνων τα οποία έχουν ενεργοποιημένες διάφορες εφαρμογές όπως GPS, Google Maps κτλ
 - Συσκευές και αισθητήρες οι οποίες είναι συνδεδεμένες στο διαδίκτυο των πραγμάτων IOT

Data Evolution

10% are structured and 90% are unstructured like emails, videos, facebook posts, website clicks etc.



Source: Hortonworks.com

Παραδείγματα

- Παραδείγματα στατιστικής ανάλυσης και εξαγωγής γνώσης (π.χ. συσχετίσεις, προβλέψεις, αναγνώριση συναισθημάτων κτλ)

- Η Gartner (η μεγαλύτερη επιχείρηση στον κόσμο που ασχολείται με την τεχνολογική έρευνα και συμβουλευτική), το 2012 έδωσε τον εξής ορισμό: «Τα big data είναι μεγάλου όγκου δεδομένα, τα οποία δημιουργούνται με «υψηλή ταχύτητα» και περιλαμβάνουν ποικιλία στοιχείων (π.χ. εικόνες, κείμενο, βίντεο κτλ)
- Για την ανάλυση των δεδομένων αυτών απαιτούνται αποδοτικές και καινοτόμες μορφές εφαρμογές επεξεργασίας και ανάλυσης.
- Στα «μεγάλα δεδομένα» συγκαταλέγονται όλες οι πληροφορίες των social media που είναι προσβάσιμες σε όλους μας και βρίσκονται στο Διαδίκτυο, δηλαδή φωτογραφίες, video και κείμενα, καθώς και όλα τα «κλειστά δεδομένα» των διαφόρων εταιριών αλλά και των κυβερνήσεων.
- Δηλαδή, η Gartner πρότεινε έναν ορισμό που περιλάμβανε τα "τρία Vs (Volume, Velocity, Variety)"

Ορισμός από την Gartner

- Τα μεγάλα δεδομένα είναι στοιχεία υψηλού όγκου, υψηλής ταχύτητας ή / και ποικιλίας που απαιτούν οικονομικά αποδοτικές, καινοτόμες μορφές επεξεργασίας πληροφοριών που επιτρέπουν βελτιωμένη διορατικότητα, λήψη αποφάσεων και αυτοματοποίηση διεργασιών.

Επεξεργασία και εξαγωγή χρήσιμων συμπερασμάτων από τα Big Data

- Συσχετίσεις μεταξύ των Μεγάλων Δεδομένων
- Χρησιμοποιούνται οι συσχετίσεις μεταξύ αντικειμένων για να γίνουν “αντιληπτές” οι προτιμήσεις -επιλογές του κάθε καταναλωτή (δημιουργία προφίλ καταναλωτή)
- Προσπαθούν δηλαδή, να προβλέψουν τις ανάγκες και τις προτιμήσεις του ώστε στη πορεία να προβούν σε προτάσεις προς τους δυνητικούς καταναλωτές για ανάγκες που ακόμη δεν έχουν γίνει αντιληπτές από τους ίδιους

Συσχετίσεις

Πελάτες που αγοράζουν ένα προϊόν (π.χ. κάποια μάρκα Whiskey) τείνουν να αγοράσουν και κάτι άλλο (π.χ. πούρα η φυστίκια η κάποιου είδους αλλαντικού)

Πελάτες που αγοράζουν κάποιο βιβλίο θα αγοράσουν και κάποιο άλλο

Πελάτες που βλέπουν μια ταινία τείνουν να δούν και κάποια άλλη

Αιτιακή σχέση

- Με τη επεξεργασία των Μεγάλων Δεδομένων και τις συσχετίσεις που μας παρέχουν, οι άνθρωποι ψάχνουν ασυναίσθητα και την αιτιακή σχέση.
- Για παράδειγμα εάν μια ασφαλιστική εταιρία με τις συσχετίσεις των δεδομένων της αναφέρει ότι τα κόκκινα αυτοκίνητα εμπλέκονται σε λιγότερα ατυχήματα, αυτομάτως θα προσπαθήσουμε να απαντήσουμε στο γιατί (π.χ έντονο χρώμα και το προσέχουν περισσότερο οι άλλοι οδηγοί, κ.α)

Παραδείγματα αξιοποίησης των μεγάλων δεδομένων- η περίπτωση της Google

- Το 2009 οι ερευνητές της Google δημοσίευσαν ένα άρθρο στο περιοδικό Nature που έδειξε ότι κατάφεραν να εντοπίσουν τη εστία γρίπης χρησιμοποιώντας αρχεία αναζήτησης και ιστορικά δεδομένα από τα Κέντρα για την ασθένεια, έλεγχο και πρόληψη (Centers for Disease Control and Prevention-CDC).

Παραδείγματα αξιοποίησης των μεγάλων δεδομένων- η περίπτωση της Google

- Τα CDC συλλέγει πληροφορίες από νοσοκομεία και κλινικές σε μεμονωμένα περιστατικά γρίπης των ασθενών, αλλά υπάρχει χρόνος καθυστέρησης από μία έως δύο εβδομάδες στην αναφορά τους
- Η ομάδα της Google δημιούργησε έναν αλγόριθμο που συνέκρινε ερωτήματα αναζήτησης με δεδομένα από τα CDC
- Πραγματοποιώντας σχεδόν μισό δισεκατομμύριο υπολογισμούς η ομάδα προσδιόρισε σαράντα πέντε όρους αναζήτησης που έχουν ισχυρή συσχέτιση με τα κρούσματα της γρίπης

Παραδείγματα Εφαρμογών

- Ανάλυση της συμπεριφοράς και των προτιμήσεων των χρηστών και ανάλογη προσαρμογή της πολιτικής προβολής και διαφήμισης

Κατανόηση και Βελτιστοποίηση Επιχειρησιακών Διαδικασιών

- Τα Μεγάλα Δεδομένα χρησιμοποιούνται επίσης για τη βελτιστοποίηση των επιχειρηματικών διαδικασιών.

- Εντοπισμός τρομοκρατών, καταζητούμενων ή/και υποψηφίων εγκληματιών, μέσω της ανίχνευσης υπόπτων τηλεπικοινωνιακών, αγοραστικών συμπεριφορών και μετακινήσεων.

Εφαρμογές

- Επεξεργασία δεδομένων από κινητές συσκευές και εξαγωγή συμπερασμάτων (π.χ. ενημέρωση για κρούσματα COVID)

Ανάγκες σε Hardware και Software

- Η σημερινή έκρηξη δεδομένων έχει ως συνέπεια την ανάγκη για γρήγορη επεξεργασία τεράστιου όγκου πληροφορίας.
- Αυτό με τις συμβατικές τεχνολογίες είναι αδύνατον να επιτευχθεί.
- Στο cloud computing ωστόσο, με την ύπαρξη εικονικά απεριόριστου hardware (on demand), μπορούμε να μοιράσουμε τα δεδομένα μας σε εκατοντάδες διαφορετικούς υπολογιστές για να επιτύχουμε γρηγορότερη επεξεργασία τους.

Περισσότερα για τους τομείς
εφαρμογής

Εμπόριο (π.χ. Ebay, Amazon)

- Μεγάλος όγκος πληροφορίας αποθηκεύεται για τους πελάτες. Με την δυνατότητα της εξόρυξης γνώσης λαμβάνουν οι επιχειρήσεις χρήσιμες πληροφορίες για τις συνήθειες και τις τάσεις.
- Κατανόηση και στόχευση πελατών

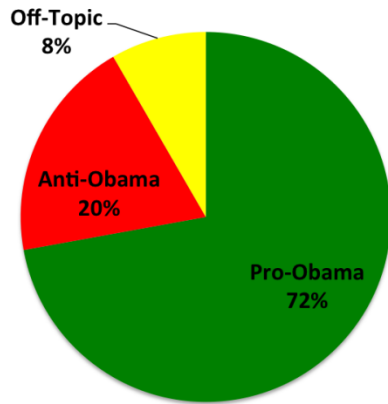
Αντιμετώπιση καταστροφών

- Μέσα από τη συγκέντρωση πληροφορίας που μπορεί να προέρχεται είτε από τις αναφορές στα κοινωνικά δίκτυα, είτε στα ΜΜΕ δίνεται η δυνατότητα χαρτογράφησης του τόπου όπου λαμβάνει χώρα μια καταστροφή, και αξιολόγησης της σοβαρότητάς.

Πολιτική

- Ανάλυση των αναφορών στο Twitter για έναν πολιτικό
- Ανάλυση συναισθήματων
- Ανάλυση ποσοτικών στοιχείων όπως re-tweets, likes κτλ

Sentiment of Tweets Mentioning Obama
4/29/2012 - 5/5/2012



Trump vs Clinton



A sentiment analysis on Twitter mentions

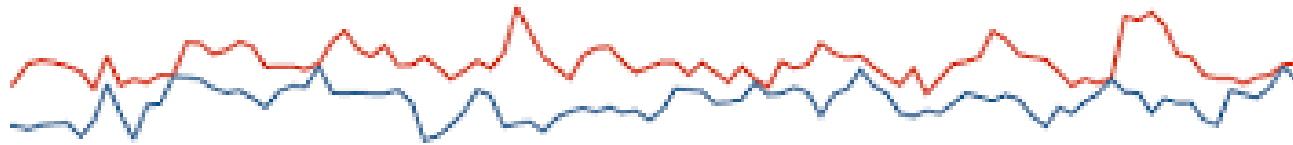




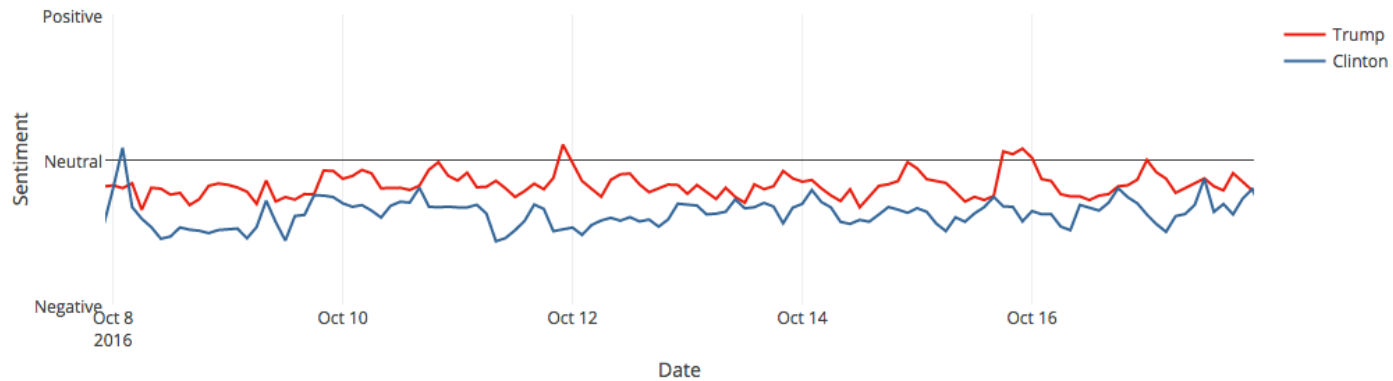
Trump vs Clinton



A sentiment analysis on Twitter mentions



Candidate sentiment over time



Φυσικοί πόροι

- Ο κλάδος του πετρελαίου θεωρείται από τους πρώτους όπου άρχισε να ασχολείται με τα μεγάλα δεδομένα.
- Πετρελαϊκές εταιρείες και κυβερνήσεις κάνουν χρήση και ανάλυση τεράστιων ποσοτήτων δεδομένων που είναι διαθέσιμα σχετικά με σεισμική δραστηριότητα σε όλη την υφήλιο με σκοπό την εξερεύνηση και εξόρυξη πετρελαίου

Υγεία

Στις μέρες μας, τα δεδομένα που αφορούν ασθενείς δημιουργούνται με εκθετικό ρυθμό. Οι πληροφορίες αυτές ωστόσο έχουν διαφορετικές μορφές και πρότυπα.

Τα πληροφοριακά συστήματα υγείας διαθέτουν μεγάλο όγκο δεδομένων

Στο μέλλον wearables (π.χ. Smart watches) και άλλες εφαρμογές θα συνεισφέρουν στο κτίσιμο του ιατρικού μας προφίλ

- Εντοπισμός μοτίβων που αφορούν στην ανταπόκριση ασθενών σε ιατρικές θεραπείες

Μεταφορές

- Η εταιρεία διεθνών ταχυμεταφορών UPS, άρχισε να καταγράφει και να μελετά τις κινήσεις πακέτων και συναλλαγών από τις αρχές της δεκαετίας 1980
- Σήμερα συγκεντρώνει δεδομένα για εκατ. πακέτα ημερησίως, και για εκατομμύρια πελάτες
- Το μεγαλύτερο τμήμα των μεγάλων δεδομένων που διαθέτει προέρχονται από τηλεματικούς αισθητήρες τοποθετημένους στα οχήματά της.
- Τα δεδομένα αξιοποιούνται τόσο για την καθημερινή εποπτεία και μέτρηση της αποδοτικότητας αλλά και για τη διαμόρφωση της βέλτιστης δομής των δρομολογίων
- Συνέπεια, της εν λόγω πρακτικής είναι η εξοικονόμηση καυσίμων

Έξυπνες πόλεις

- Τα Μεγάλα Δεδομένα χρησιμοποιούνται για τη βελτίωση πολλών πτυχών των πόλεων και χωρών.
- Για παράδειγμα , η χρήση τους επιτρέπει σε πόλεις τη βελτίωση της ροής της κυκλοφορίας.
- Κάτι τέτοιο στηρίζεται σε δεδομένα για τη κυκλοφορία σε πραγματικό χρόνο, σε δεδομένα των μέσων κοινωνικής δικτύωσης καθώς και σε κλιματολογικές πληροφορίες.
- Πολλές πόλεις αξιοποιούν πλέον τα Μεγάλα Δεδομένα και αποκτούν το τίτλο των έξυπνων πόλεων (smart cities)

Ανοιχτά Δεδομένα (open data)
και Μεγάλα Δεδομένα

Ανοιχτά δεδομένα

- Τα ανοικτά δεδομένα είναι δημόσια προσβάσιμα δεδομένα που οι άνθρωποι, οι επιχειρήσεις και οι οργανισμοί μπορούν να χρησιμοποιούν για την έναρξη νέων επιχειρήσεων, την ανάλυση προτύπων και τάσεων, να λαμβάνουν αποφάσεις που βασίζονται σε δεδομένα, και να επιλύουν σύνθετα προβλήματα

Χαρακτηριστικά

- Τα δεδομένα πρέπει να είναι διαθέσιμα στο κοινό και σε οποιονδήποτε θελήσει να τα χρησιμοποιήσει
- Πρέπει να διαθέτουν άδεια κατά τέτοιο τρόπο που να επιτρέπει την επαναχρησιμοποίησή τους.

Παραδείγματα- Συζήτηση

- Worldometer
- <https://www.worldometers.info/sources/>