



Ανοικτά Δεδομένα

Προκλήσεις και Ευκαιρίες:

Η εμπειρία του OpenDataCloud

Κώστας Σαΐδης, PhD

Πάροχοι Ανοικτών Δεδομένων

- datagon.gr
- diavgeia.gr
- geodata.gov.gr
- Πυροσβεστικό σώμα
- Ελληνική Αστυνομία
- Υπουργείο Ανάπτυξης
- Ελληνική Στατιστική Αρχή
- κλπ

Το (κύριο) Πρόβλημα

- Κάθε εφαρμογή που θέλει να αξιοποιήσει τα ανοικτά δημόσια δεδομένα, θα πρέπει να αντιμετωπίσει μια σειρά κοινών ζητημάτων

Για παράδειγμα

- Πού και πώς αποθηκεύω τα δεδομένα;
 - Πού και πώς τα ευρετηριάζω;
 - Πώς τα ενσωματώνω στην εφαρμογή μου;
-
- Ο καθένας τα λύνει μόνος του, με ad-hoc τρόπο!

Ναι, αλλά...

- Αν προκύψει:
 - μια νέα έκδοση των δεδομένων;
 - ένα νέο σύνολο δεδομένων;
 - μια νέα εφαρμογή;
- Ξανά από την αρχή...!

Το OpenDataCloud

- Μια υπηρεσία νέφους (cloud service) για την αξιοποίηση των ανοικτών δημόσιων δεδομένων.

Μια πλατφόρμα

Που λύνει τα κοινά θέματα

διαχείρισης, προσπέλασης, αναζήτησης των δεδομένων

Σε κεντρικό επίπεδο

χρησιμοποιώντας τεχνολογίες αιχμής, βέλτιστες
πρακτικές διαλειτουργικότητας, ανοικτά πρότυπα

Επιτρέποντας την ενσωμάτωση των δεδομένων

σε οποιαδήποτε εφαρμογή με αποτελεσματικό τρόπο

Οφέλη

- Κοινές βάσεις που επιτρέπουν σε όλους, να ενσωματώσουν τα ανοικτά δεδομένα:
 - σε οποιοδήποτε σενάριο ή εφαρμογή,
 - με εύκολο και αποτελεσματικό τρόπο
- Προστιθέμενη Αξία:
 - Ταχύτητα, απλότητα, αυτοματισμός, οικονομία κλίμακας, συμφωνία με τις βέλτιστες διεθνείς πρακτικές

Παρουσίαση του OpenDataCloud

www.opendatacloud.gr

OpenDataCloud Software Stack

- Server Side
 - Java / Groovy
 - Restlet
 - ElasticSearch 1.x
- Browser Side
 - Bootstrap
 - D3.js
 - Knockout.js

To software stack του OpenDataCloud
είναι 100% open source.



open source

Open Data

- Open data ≠ File sharing
- Identifiable, navigable and searchable data records
 - Linked data & SPARQL/RDF
 - NoSQL (OpenDataCloud)

Open Data

- “A piece of data is open if anyone is free to **use, reuse, and redistribute** it — subject only, at most, to the requirement to attribute and/or share-alike.”

Data use, reuse, redistribute

- Data querying, filtering & transformation
- Data embedding & integration
- Data exchange & interoperability
- Data enhancement & evolution
- Data validity & accountability
- ...

Data-centric Οργανισμοί

- Ας δούμε το δημόσιο ως έναν οργανισμό (π.χ. μια επιχείρηση)
- Με διαφορετικά τμήματα διάσπαρτα σε διαφορετικά γεωγραφικά σημεία
- Που χρησιμοποιούν **διαφορετικά λογισμικά για διαφορετικά δεδομένα**

Συμπέρασμα

- Οι προκλήσεις των ανοικτών δεδομένων
- Σε μικρότερη ή μεγαλύτερη κλίμακα
- Απαντώνται σε όλους τους σύγχρονους οργανισμούς

Software & Data

- Το λογισμικό που χρησιμοποιούμε περιορίζει την ανοικτότητα των δεδομένων μας (δηλαδή το ποσοστό της ελευθερίας που έχουμε πάνω στα δεδομένα)
- E.g. vendor lock-in, legacy applications, etc.

Βέλτιστες πρακτικές για data openness

Από την εμπειρία του OpenDataCloud:

1. Data indexing/searching
2. RESTful Web APIs

Ο στόχος

Εντός του οργανισμού:

- η βασική διαχείριση (create, update, delete & search) όλων των δεδομένων
- να γίνεται μέσω RESTful Web-based APIs.

Τι είναι το REST

- Ένα αρχιτεκτονικό στυλ για τη λειτουργία του Παγκόσμιου Ιστού
 - Οριοθετεί αρχές, περιορισμούς και βασικές λειτουργίες
 - Ανεξάρτητα της γλώσσας προγραμματισμού, του πρωτοκόλλου επικοινωνίας ή του είδους των δεδομένων

ΣΥΝΟΠΤΙΚΑ

- Τέσσερις βασικές έννοιες:
 - Resources, Representations, Requests, Responses
- Έξι βασικές αρχές:
 - Client-server, Stateless, Cacheable, Layered System, Uniform Interface, Code on demand (προαιρετικά)

Οφέλη

- Η σωστή χρήση των RESTful αρχών βελτιώνει όλα τα σημαντικά χαρακτηριστικά μιας αρχιτεκτονικής:
 - Απόδοση
 - Κλιμάκωση
 - Απλότητα
 - Επεκτασιμότητα
 - Αξιοπιστία

Best Practice (1)

- Παροχή RESTful APIs για όλες τις πηγές δεδομένων του οργανισμού

ElasticSearch

- A search engine:
 - For searching and analyzing data in real time
 - That is distributed, scalable, and highly available
 - And offers a powerful RESTful API
- Open Source (Apache License)
- Based on Apache Lucene
- Supported by elastic (the company)

Best practice (2)

- Χρήση του Elasticsearch για τη συγκέντρωση όλων των δεδομένων από τις πιθανές άλλες πηγές εντός του οργανισμού σε ένα σημείο
- Ευρετηρίαση και αναζήτηση των δεδομένων μέσω Web APIs
- Clustered, Distributed, Scalable, Rebuildable

OpenDataCloud

- Ο (μάλλον φιλόδοξος) στόχος:
 - Συγκέντρωση όλων των ανοικτών & δημόσιων πηγών δεδομένων
 - Πλήρης αυτοματοποίηση της ροής εργασίας «δημοσίευσης» των δεδομένων
 - Αυτόματη ενημέρωση όλων των εφαρμογών

Ανοικτά Ζητήματα

- Καθιέρωση
 - μοναδικών προσδιοριστών για τις πηγές
 - ενιαίας και συνεπούς διαδικασίας δημοσίευσης και ενημέρωσης των πηγών
 - ενιαίας περιγραφής (metadata) των πηγών
 - αδειών χρήσης για τα δεδομένα των πηγών
 - βέλτιστων πρακτικών «δόμησης» των πηγών

ΕΠΕΚΤΑΣΕΙΣ

- Προσθήκη νέων πηγών και γραφημάτων
- Επέκταση του API – Embedding σε σελίδες με απλό copy/paste
- Πρόσθετες δυνατότητες αναζήτησης
- Υποστήριξη αγγλικών
- RSS Feeds με ενημερώσεις για νέες πηγές

Κοινές Δράσεις & Συνεργασίες

- Σύμπραξη με την ευρύτερη κοινότητα των Ανοικτών Δεδομένων
- Community-driven curation για τα DataSets
- Υλοποίηση πρωτοκόλλου για τις πηγές δεδομένων
- Συγγραφή οδηγιών για τη δόμηση των πηγών
- Κατασκευή πρόσθετων εφαρμογών

Σας Ευχαριστώ 😊

saiko@niovity.com

