

Ανοικτά Δεδομένα

Ρούτσης Φώτιος

Αθήνα, 26-10-2016

Attribution-ShareAlike 4.0 International



ΡΟΥΤΣΗΣ ΦΩΤΙΟΣ



fotis@routsis.gr



Οργανισμός Ανοικτών Τεχνολογιών ΕΕΛΛΑΚ



- Ιδρύθηκε το **2008**
- **Μη κερδοσκοπικός Οργανισμός**
- Μέλη: 30 **Ακαδημαϊκά** Ιδρύματα,
Ερευνητικά Κέντρα και
Κοινωφελείς Οργανισμοί



Οργανισμός Ανοικτών Τεχνολογιών ΕΕΛΛΑΚ



- Εκπαίδευση για τα ανοιχτά δεδομένα
- Εκδηλώσεις
- **Πιστοποίηση** εκπαιδευτών & δεδομένων
- Διεθνείς συνεργασίες: **Open Data Institute**



Open Data institute

Ιδρύθηκε από

- *Sir Tim Berners-Lee*
- *Sir Nigel Shadbolt*

Inventor of the WWW

Non Profit

65 Clifton Street, London, UK



“... to catalyse the evolution of open data culture and to create economic, environmental, and social value.”



WHY OPEN DATA?

THE DATA ECONOMY



by <http://raconteur.net/>

Annual turnover
€130bn
500K employees

**270 open data
companies
in the UK**



Open Data Institute (2015) Open data means business: UK innovation across sectors and regions. London, UK. Available at <http://theodi.org/open-data-means-business-uk-innovation-sectors-regions>

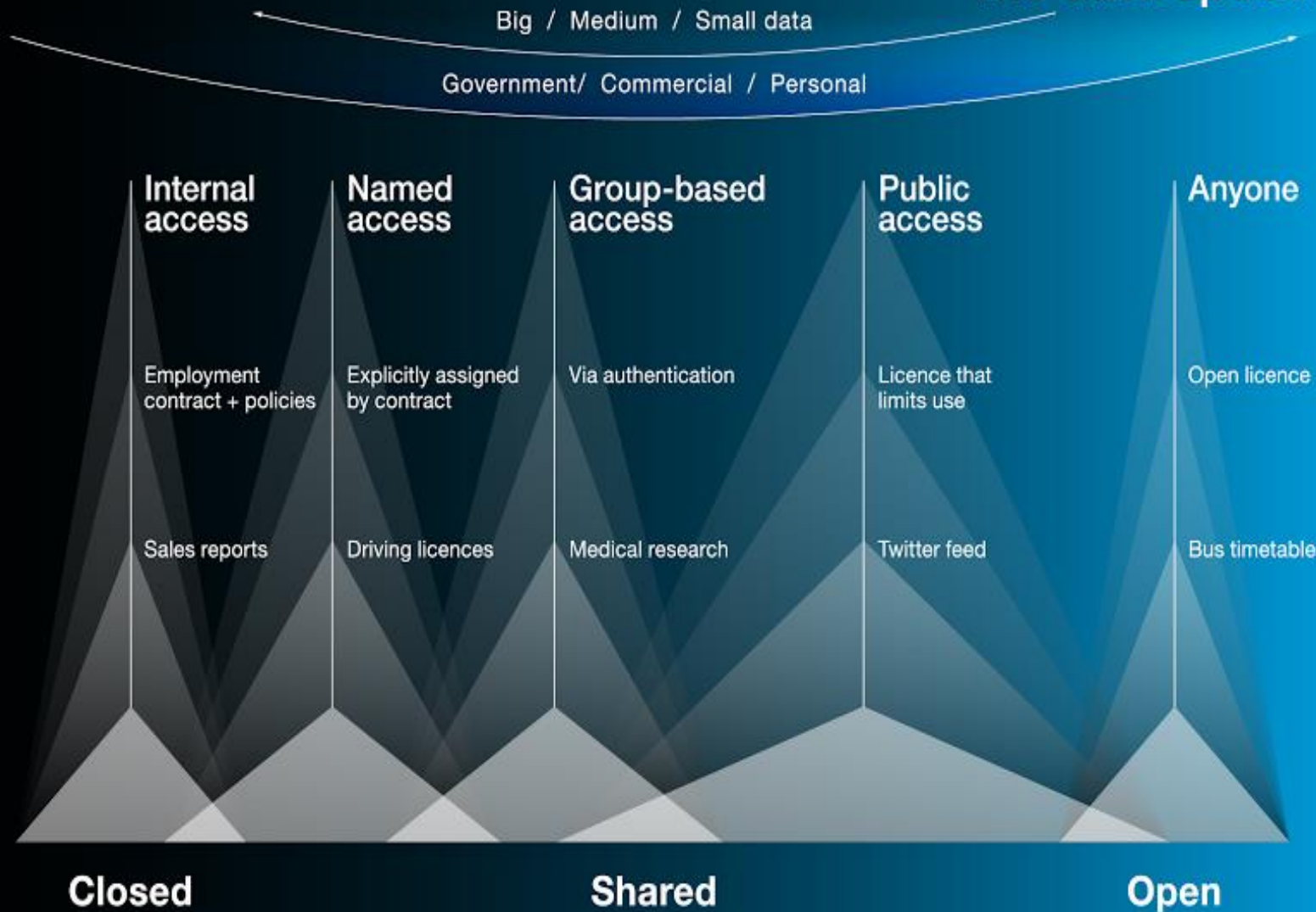
Ορισμός

Ανοικτά Δεδομένα είναι τα Δεδομένα που μπορεί οποιοσδήποτε να έχει **πρόσβαση**, να τα **χρησιμοποιεί** και να τα **διανέμει**.

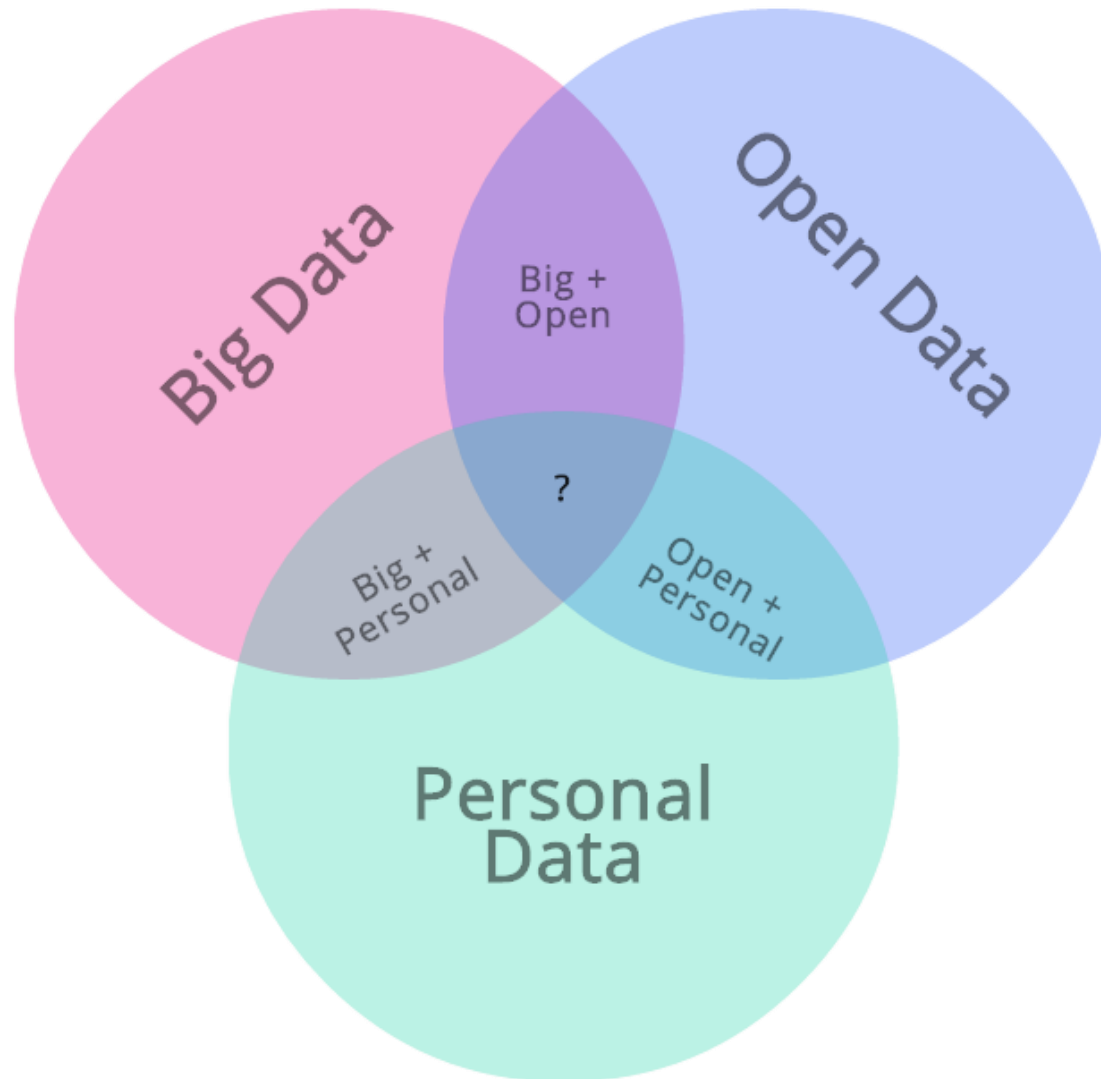
- Open Data Institute
- <http://theodi.org/guide/what-open-data>

Ορισμός από το 2015

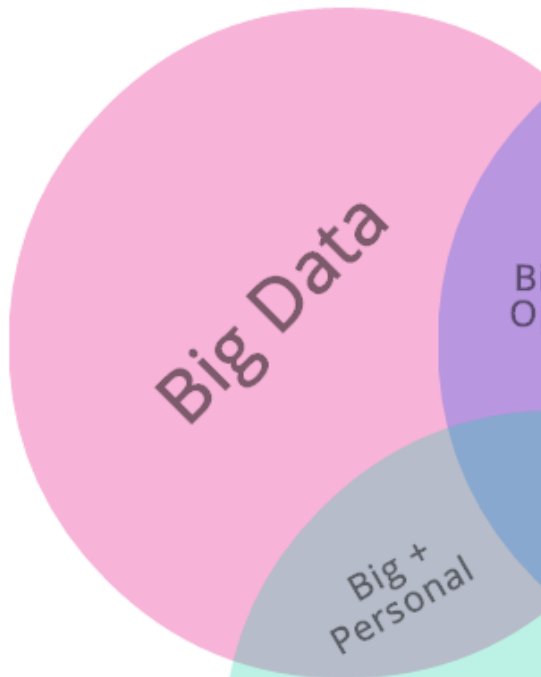
The data spectrum



Που κινούμαι ;

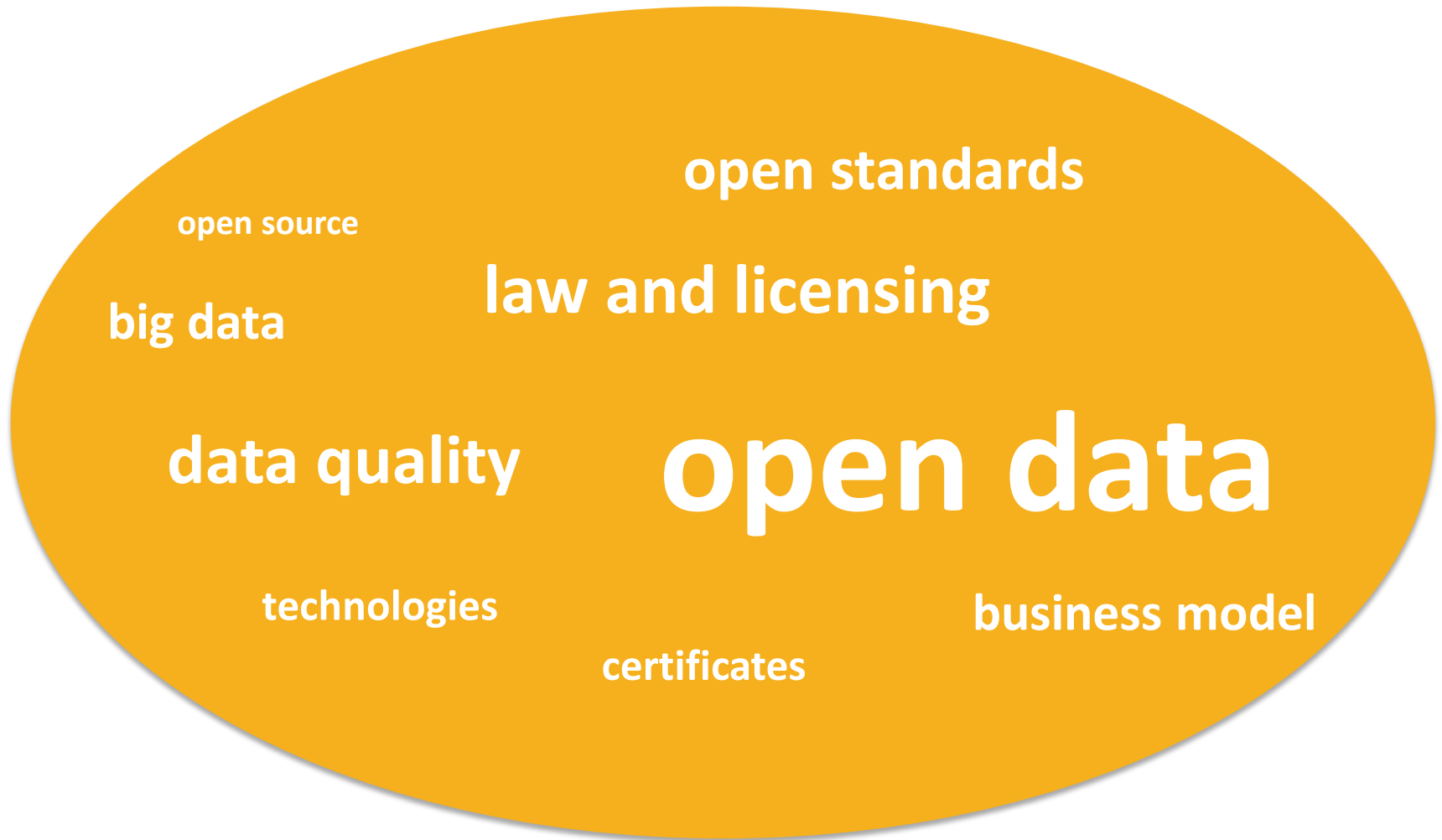


Big Data



Χρειάζονται εξειδικευμένα
εργαλεία για να τα
επεξεργαστούμε

Τα έμαθα όλα ;



Σημαντικό!

Τα Ανοικτά Δεδομένα που
χρησιμοποιούμε θα πρέπει να φέρουν
κατάλληλη **άδεια χρήσης**

Creative Commons

Public Domain

Open Database Licence

EUPL ..



ΠΟΥ ΤΑ ΒΡΙΣΚΩ ;

Αναζήτηση..

Τα Ανοικτά Δεδομένα παρέχονται από πολλούς διαφορετικούς οργανισμούς, οπότε δεν υπάρχει μοναδικό σημείο όπου μπορούν να εντοπιστούν/αναζητηθούν.

Μεθοδολογία..

- Αποφασίστε τι είδους δεδομένα αναζητάτε
- Εντοπίστε τους οργανισμούς οι οποίοι είναι υπεύθυνοι γι' αυτά
- Επισκεφθείτε την ιστοσελίδα τους..

<https://github.com/theodi/shared/wiki/Finding-Open-Data>

Μεθοδολογία..

- Λέξεις κλειδιά για την αναζήτηση δεδομένων:
 - data / δεδομένα
 - research / έρευνα
 - reports / αναφορές
 - publications / δημοσιεύσεις
 - transparency / διαφάνεια

<https://github.com/theodi/shared/wiki/Finding-Open-Data>

Πηγές Δεδομένων | Gov

Εθνικά

- data.gov.gr
- geodata.gov.gr
- diavgeia.gov.gr

Διεθνώς

- data.gov.uk
- data.gov
- ...

EU

- open-data.europa.eu/en/data/

Πηγές Δεδομένων | Community

Εθνικά

- go.eellak.gr/datasets
- ckan.linkedeconomy.org/dataset

Διεθνώς

- github.com/theodi/shared/wiki/Finding-Open-Data



ΕΡΓΑΛΕΙΑ ;

Ναι!

Υπάρχουν εργαλεία για όλες τις ανάγκες:

Ανάγνωση

Οπτικοποίηση

Ανάλυση

Επεξεργασία & Καθαρισμός

Δημοσίευση

Οπτικοποίηση

Google Fusion Tables (The Alpha)

http://tables.googlelabs.com/DataSource?table=00000000

benjaminmiller@bath.ac.uk | Log In | FAQ | Feedback | Sign Out

Total Renewable Freshwater Supply, by Country (2006 Update)

Pacific Institute (http://www.pacificinst.org)

File View Edit Visualize Merge

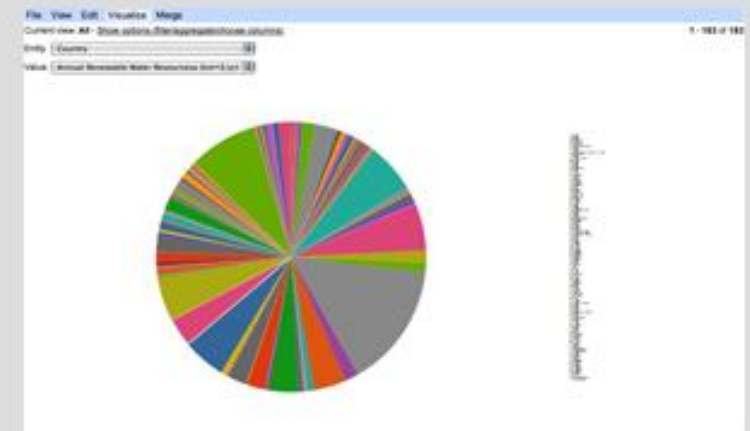
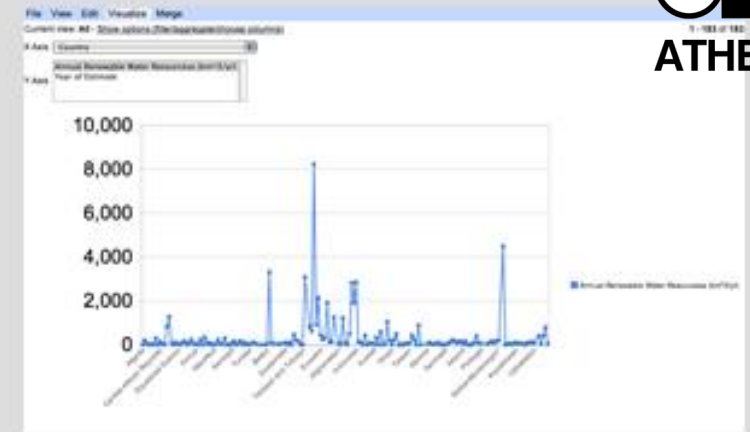
Current view: All - Show options, Filter by geographic column(s)

1 - 100 of 183 Rows

Country	Annual Renewable Water Resource	Year of Estimate	Location
Algeria	18.2	1987	Estimates from FAO (1985) "Water Resources of the ..."
Angola	184.0	1987	Estimates from Benjamen, Institute of Geography, US...
Bahrain	20.8	2001	Wargat, J.COS, 2001. Gas resources on the ...
Botswana	14.7	2001	Wargat, J.COS, 2001. Gas resources on the ...
Burkina Faso	17.6	2001	Wargat, J.COS, 2001. Gas resources on the ...
Burundi	3.6	1987	Estimates from Benjamen, Institute of Geography, US...
Cameroon	269.5	2003	Estimates from FAO (2003) "Review of World Water R...
Cape Verde	8.3	1980	Estimates from FAO (1980) "Water Resources of the ..."
Central African Republic	164.4	2003	Estimates from FAO (2003) "Review of World Water R...
Chad	43.0	1987	Estimates from Benjamen, Institute of Geography, US...
Comoros	1.2	2003	Estimates from FAO (2003) "Review of World Water R...
Congo	832.0	1987	Estimates from Benjamen, Institute of Geography, US...
Congo, Democratic Republic (formerly Zaire)	1283	2001	Wargat, J.COS, 2001. Gas resources on the ...
Cote D'Ivoire	81	2001	Wargat, J.COS, 2001. Gas resources on the ...
Cyprus	0.3	1987	Estimates from FAO (1987) "Water Resources of the ..."
Egypt	86.8	1987	Estimates from FAO (1987) "Water Resources of the ..."
Equatorial Guinea	26	2001	Wargat, J.COS, 2001. Gas resources on the ...
Eritrea	8.3	2001	Wargat, J.COS, 2001. Gas resources on the ...
Ethiopia	110.0	1987	Estimates from Benjamen, Institute of Geography, US...
Gabon	154.0	1987	Estimates from Benjamen, Institute of Geography, US...
Gambia	9.0	1982	Estimates from FAO (1982) "Water Resources of the ..."
Ghana	63.2	2001	Wargat, J.COS, 2001. Gas resources on the ...
Guinea	226.0	1987	Estimates from Benjamen, Institute of Geography, US...
Guinea-Bissau	31.0	2003	Estimates from FAO (2003) "Review of World Water R...
Kampuchea	36.3	1980	Estimates from FAO (1980) "Water Resources of the ..."
Lesotho	9.2	1987	Estimates from Benjamen, Institute of Geography, US...
Libya	231.0	1987	Estimates from Benjamen, Institute of Geography, US...

Google Fusion Tables

tables.googlelabs.com



Οπτικοποίηση

MAP YOUR WORLD'S DATA

CartoDB is the easiest way to map and analyze your location data

SIGN UP

<https://cartodb.com/>

Επεξεργασία / Καθαρισμός



ATHENS

Custom text transform on column Column

Expression

Language

Google Refine Expression Language (GREL)

```
value.replace("''", "")
```

yntax error.

Refine



Preview

History

Help

row value

1. '''1954 [[Meena Kumari]] - "[[Baiju Bawra (1952 film)|Baiju Bawra]]" as "Gauri"

2. '''1955 [[Meena Kumari]] - "[[Parineeta (1953 film)|Parineeta]]" as "Lolita"

3. '''1956 [[Kamini Kaushal]] - "[[Biraj Bahu]]" as "Biraj Chakravorty"

4. '''[[Geeta Bali]] - "[[Vachan]]" as Kamla

5. '''[[Meena Kumari]] - "[[Azaad (1955 film)|Azaad]]" as Shobha

6. '''1957 [[Nutan]] - "[[Seema (1955 film)|Seema]]" as Gauri

value.replace("''", "")

*1954 [[Meena Kumari]] - "[[Baiju Bawra (1952 film)|Baiju Bawra]]" as Gauri

*1955 [[Meena Kumari]] - "[[Parineeta (1953 film)|Parineeta]]" as Lolita

*1956 [[Kamini Kaushal]] - "[[Biraj Bahu]]" as Biraj Chakravorty

*[[[Geeta Bali]] - "[[Vachan]]" as Kamla

*[[[Meena Kumari]] - "[[Azaad (1955 film)|Azaad]]" as Shobha

*1957 [[Nutan]] - "[[Seema (1955 film)|Seema]]" as Gauri

On error ☒ set to blank

☐ store error

☐ keep original

☐ Re-transform up to 10 times until no change

OK

Cancel

<http://openrefine.org/>

Επεξεργασία / Καθαρισμός

DataCleaner 4.5^{new!}
The premier data quality solution

(commercial) open source

<http://datacleaner.org/>

Κάτι πιο απλό;



MARKETING SPEND BY MONTH



Web Services;

Διατίθενται πληθώρα ΕΛ/ΛΑΚ εργαλείων για την **δημοσίευση** των δεδομένων μας μέσω Web Services.



https://en.wikipedia.org/wiki/List_of_web_service_frameworks

Άλλες προτάσεις ;

- WEKA
- SPSS
- Tableau Public
- Many Eyes
- OpenLayers
- OpenStreetMaps
-



ΕΧΩ ΤΑ ΔΕΔΟΜΕΝΑ.. ΤΩΡΑ ;

Ποιότητα Δεδομένων

Μια από τις σημαντικότερες ιδιότητες που προσδιορίζουν την ποιότητα των δεδομένων που παρέχουμε είναι η **ορθότητά** τους και ο **μονοσήμαντος προσδιορισμός** τους..



Γιατί καθαρά δεδομένα ;

Έρευνες έχουν δείξει ότι ο καθαρισμός και προετοιμασία των δεδομένων για χρήση από τον καταναλωτή μπορεί να καταλάβει μέχρι και το **80%** του συνολικού χρόνου !

Πλήρη δεδομένα

- Ελλιπείς σειρές δεδομένων
- **Άδεια κελιά**

Τα δεδομένα θα πρέπει να είναι όσο το δυνατόν πιο πλήρη.

Μονοσήμαντη Αναπαράσταση

- **Ημερομηνίες**

- 22/10/2015
- 10/22/2015
- 2015/10/22

- **Σύμβολα**

- \$ € £ ¥
- C, F

Τα δεδομένα θα πρέπει να έχουν μονοσήμαντη προσέγγιση στην αναπαράσταση των δεδομένων

Αριθμοί

- Δεκαδικοί,
- Χιλιάδες
 - 1000.23
 - 1000,23
 - 1.000,23
 - 1,000.23

- Αθροίσματα

- $A + B = C$
- $A + B + C ?!$

Αποφεύγουμε
συγκεντρωτικά
δεδομένα..

Συμβολοσειρές

- **Ακρωνύμια**
 - UN, OHE
 - UK, HB
- **Αναπαραστάσεις**

Προσεκτική χρήση
συμβολοσειρών.

Χρήση λεξικού.

Περισσότερα..

- **Διπλές εγγραφές**
- **Επαναλαμβανόμενες εγγραφές**
- **Δεδομένα επεξηγηματικά**
- **Κεφαλίδες/Ονομασίες κελιών**



ΜΟΡΦΟΤΥΠΑ

Μορφότυπος

Σαφώς προσδιορισμένο και
προτυποποιημένο τρόπο κωδικοποίησης
ψηφιακής πληροφορίας (για αποστολή ή για
αποθήκευση σε αρχείο), έτσι ώστε ο
παραλήπτης του μηνύματος (ενήμεροι για το
χρησιμοποιούμενο μορφότυπο) να μπορούν
να εξάγουν την ορθή πληροφορία από το
μήνυμα ή από το αρχείο.

Comma Seperated Values

CSV

```

1 36301122334,Message text 1,4/29/2010 12:30
2 36301122335,Message text 2,4/30/2010 12:30
3 36301122336,Message text 3,5/1/2010 12:30
4 36301122337,Message text 4,5/2/2010 12:30
5 36301122338,Message text 5,5/3/2010 12:30
6 36301122339,Message text 6,5/4/2010 12:30
7 36301122340,Message text 7,5/5/2010 12:30
8 36301122341,Message text 8,5/6/2010 12:30
9 36301122342,Message text 9,5/7/2010 12:30
10 36301122343,Message text 10,5/8/2010 12:30
11 36301122344,Message text 11,5/9/2010 12:30
12 36301122345,Message text 12,5/10/2010 12:30
13 36301122346,Message text 13,5/11/2010 12:30
14 36301122347,Message text 14,5/12/2010 12:30
15 36301122348,Message text 15,5/13/2010 12:30
16 36301122349,Message text 16,5/14/2010 12:30
17 36301122350,Message text 17,5/15/2010 12:30
18 36301122351,Message text 18,5/16/2010 12:30
19 36301122352,Message text 19,5/17/2010 12:30
20 36301122353,Message text 20,5/18/2010 12:30
  
```

eXtensible Markup Language

XML

```

<SampleXML>
  <Colors>
    <Color1>White</Color1>
    <Color2>Blue</Color2>
    <Color3>Black</Color3>
    <Color4 Special="Light">Green</Color4>
    <Color5>Red</Color5>
  </Colors>
  <Fruits>
    <Fruits1>Apple</Fruits1>
    <Fruits2>Pineapple</Fruits2>
    <Fruits3>Grapes</Fruits3>
    <Fruits4>Melon</Fruits4>
  </Fruits>
</SampleXML>
  
```

JavaScript Object Notation

JSON

```
{
  hey: "guy",
  anumber: 243,
  - anobject: {
    whoa: "nuts",
    - anarray: [
      1,
      2,
      "thr<h1>ee"
    ],
    more: "stuff"
  },
  awesome: true,
  bogus: false,
  meaning: null,
  japanese: "明日がある。",
  link: http://jsonview.com,
  notLink: "http://jsonview.com is great"
}
```

Resource Description Framework

RDF

SPARQL Queryable



Γεωγραφικά Δεδομένα Γεωχωρικά Δεδομένα 3διάστατες αναπαραστάσεις

KML

<http://www.opengeospatial.org/standards/kml>

SHP

Shapefile

5-Stars



<http://5stardata.info/>



OL RE



OL



OL RE OF

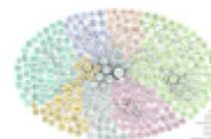


OL RE OF URI

<http://data...>



OL RE OF URI LD



OPEN DATA



5 stars



Διαθέσιμα στο διαδίκτυο σε οποιαδήποτε μορφή αλλά με Άδεια χρήσης Ανοικτών Δεδομένων



Διαθέσιμα σε μηχαναγνώσιμη μορφή (πχ excel αντί για σκαναρισμένο έγγραφο)



Χρήση ελεύθερου μορφότυπου (CSV αντί για excel)



χρήση URIs για την επισήμανση των δεδομένων μας (με τα URI να δίνουν πληροφορίες για τα δεδομένα μας)



Όλα τα παραπάνω συν τη διασύνδεση των δεδομένων μας με άλλα δεδομένα (linked data).



Open mHealth

Παράγει **πρότυπα** και **βιβλιοθήκες** για IoT και mHealth λύσεις.

Open Source: **Apache 2.0 Licence**

Publicly available: **Github**

openmhealth.org



Blood glucose (example)

Sample Data

< With temporal relation... >

```
1  {
2    "blood_glucose": {
3      "unit": "mg/dL",
4      "value": 95
5    },
6    "effective_time_frame": {
7      "date_time": "2013-02-05T20:25:00Z"
8    },
9    "temporal_relationship_to_sleep": "before sleeping"
10 }
```



Blood glucose (example)

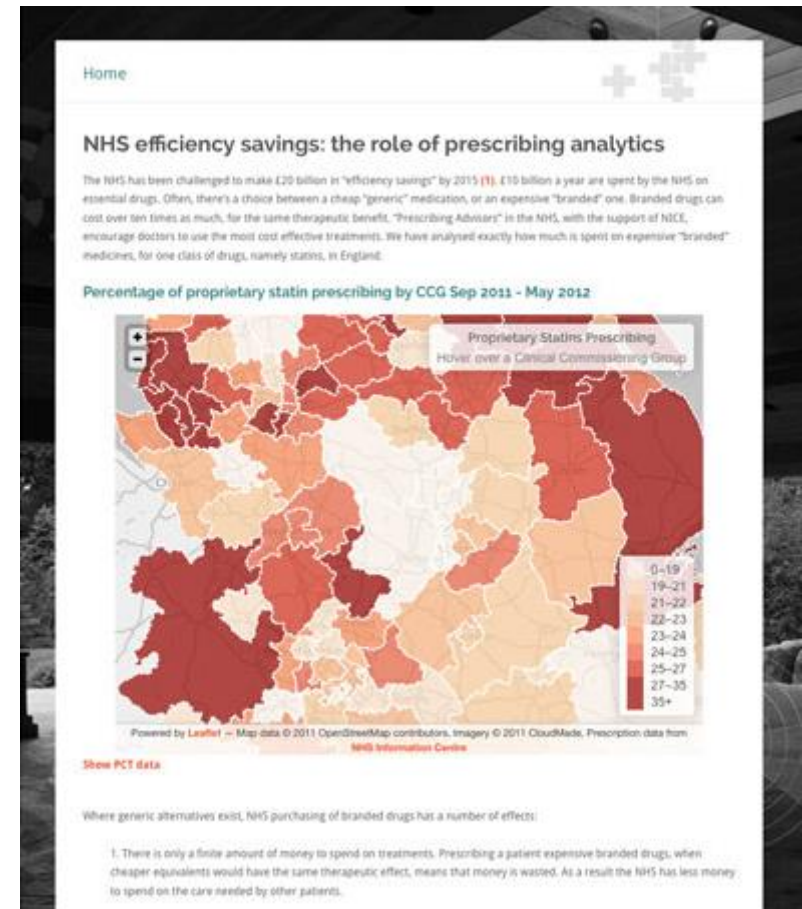
```
"properties": {
  "blood_glucose": {
    "allOf": [
      {
        "$ref": "#/definitions/unit_value"
      },
      {
        "properties": {
          "unit": {
            "enum": [
              "mg/dL",
              "mmol/L"
            ]
          }
        }
      }
    ]
  }
},
]
},
"specimen_source": {
  "allOf": [
    {
      "$ref": "#/definitions/specimen_source"
    },
    {
      "enum": [
        "interstitial fluid",
        "capillary blood",
        "plasma",
        "serum",
        "tears"
      ]
    }
  ]
}
```



CASE STUDIES

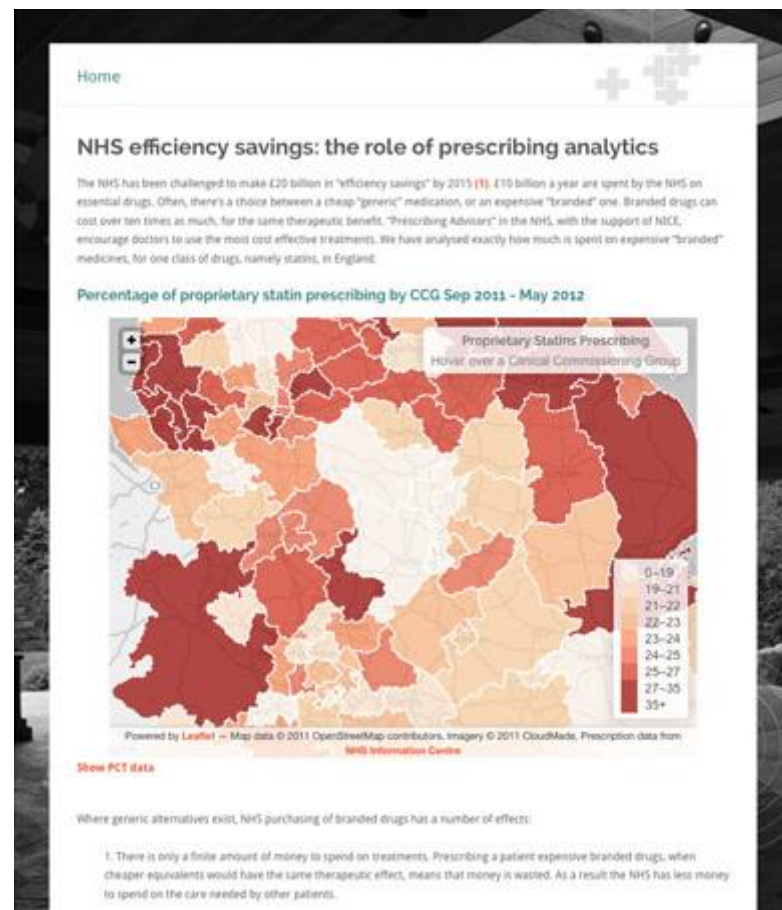
Ποιος ;

- Mastodon C
- Open Health Care UK



Δεδομένα ;

- NHS Information Centre
- 11 εκατομμύρια
εγγραφές
συνταγών



Σκοπός ;

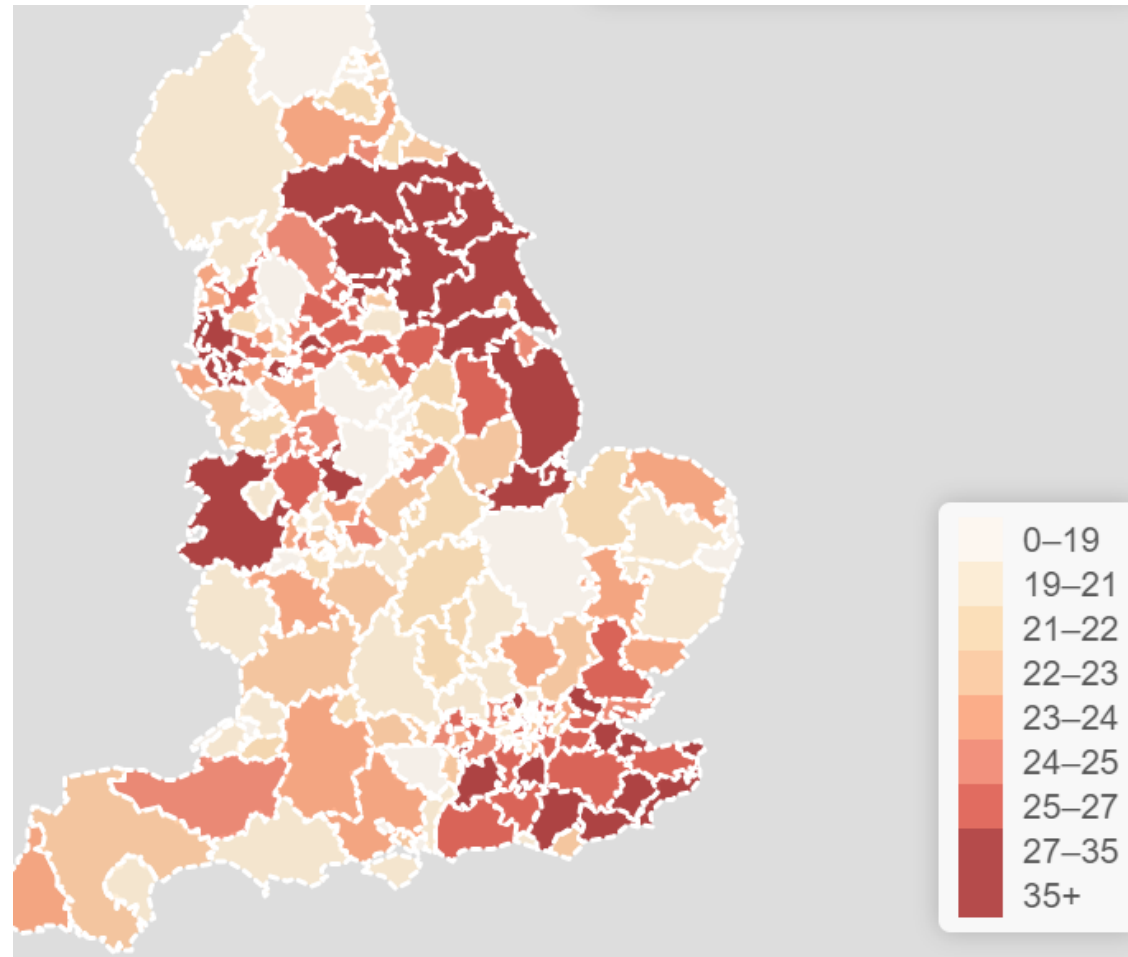
Θέλησαν να εντοπίσουν την αποτελεσματική η μη συνταγολόγηση φαρμάκων στατινών (statins) στο ΗΒ.

Γενόσημα VS Branded

Αποτέλεσμα ;

£200m θα
μπορούσαν
να «σωθούν»

..



Διαδικασία ;

Η διαδικασία περιελάμβανε

- τις συνταγές ανά χημική ουσία από το Σεπτ. 11
- Τα γεωγραφικά διαμερίσματα του ΗΒ
- Μαθηματική αναγωγή για την εκτίμηση του κόστους των φαρμάκων.



Εύκολο ;

Κάθε μηνιαίο αρχείο με τις συνταγές που χρησιμοποίησαν ήταν τύπου **.CSV** και ήταν μεγέθους περίπου **500MB** ..

<https://github.com/MastodonC/nhs-prescription-analytics>

Open Data VS Ebola

- Το 2014 ξεκίνησε μια από τις μεγαλύτερες επιδημικές εξάρσεις του ιού Ebola
- Η Σιέρρα Λεόνε δέχθηκε το μεγαλύτερο πλήγμα
- Για την αντιμετώπιση της κρίσης αξιοποιήθηκαν τα ανοικτά δεδομένα

Open Data VS Ebola



Search Datasets

DATA | LOCATIONS | ORGANISATIONS | CRISES ▼ | FAQ

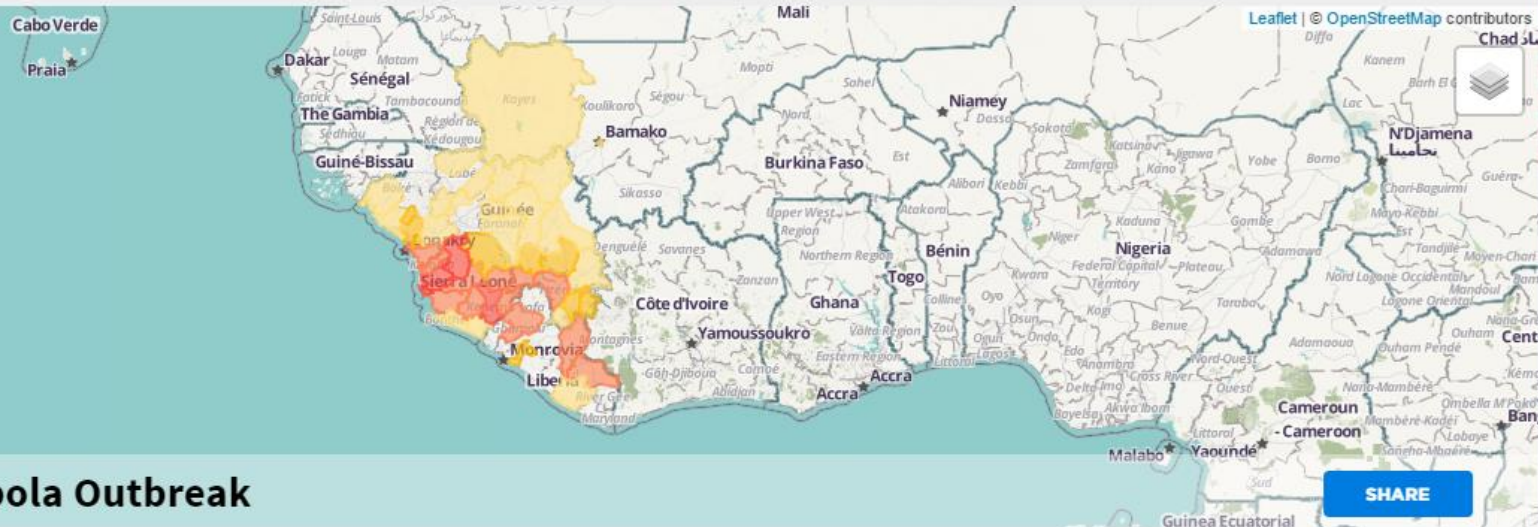
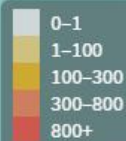
ADD DATA

HOME / EBOLA



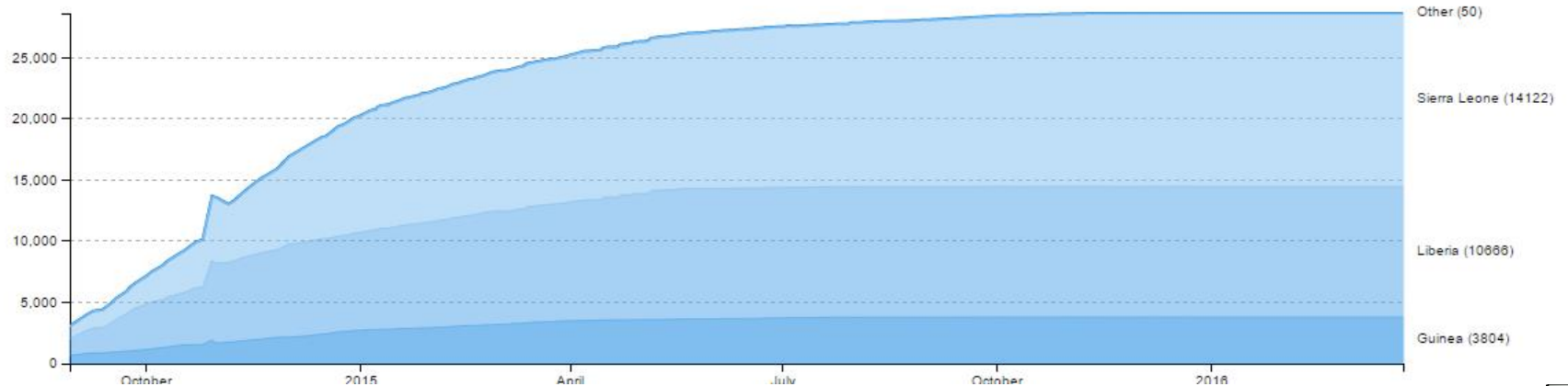
Cumulative Cases of Ebola

Hover over a country/district



West Africa: Ebola Outbreak

Week by Week Spread of Ebola in West Africa



Homework

- Θέλω για μια εφαρμογή κινητού να έχω τις συντεταγμένες (χάρτης) των Νοσοκομείων και τα στοιχεία επικοινωνίας τους.

Ερωτήματα

- Πόσα είναι τα Ελληνικά Νοσοκομεία;
- Υπάρχει έτοιμη η πληροφορία αυτή;
- Που μπορώ να βρώ ;

Homework

Δι@ύγεια API

<https://diavgeia.gov.gr/luminapi/api/organizations/supervised/all/100010899?withDecisions=true>

Επιστρέφει όλους τους εποπτευόμενους φορείς του Υπ. Υγείας.

Homework

Δι@ύγεια API

1. Απομονώνουμε όσα έχουν την λέξη Νοσοκομείο στον τίτλο τους
2. Διαχείριση μέσω Google Fusion Tables και αποτύπωση σε έναν χάρτη.
3. Εμφανή προβλήματα ;

ΕΥΧΑΡΙΣΤΩ

