Iaac | Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA
2021/2022

FACULTY DIEGO PAJARITO

Big Data

Where is all that oil?

2017

Regulating the internet giants

The world's most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

📖 Print edition | Leaders ›
May 6th 2017

A NEW commodity spawns a lucrative, fast-growing industry, prompting antitrust regulators to step in to restrain those who control its flow. A century ago, the resource in question was oil. Now similar concerns are being raised by the giants that deal in data, the oil of the digital era. These titans—Alphabet (Google's parent company), Amazon, Apple, Facebook and Microsoft—look unstoppable. They are the five most valuable listed firms in the world. Their profits are surging; they collectively racked up over

https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

2019

Here's Why Data Is Not The New Oil

Bernard Marr Contributor ⓘ
Enterprise & Cloud

It's a claim you've probably heard multiple times – "Data is the new oil!"

Now it's true, that in some ways, the analogy fits – it's easy to draw parallels due to the way information (data) is used to power much of the transformative technology we see today – artificial intelligence, automation and advanced, predictive analytics.

Adobe Stock   ADOBE STOCK

However, in many ways, it's also lazy and inaccurate – and while it's handy as a marketing shorthand (because it gets across the fact that data is a valuable commodity with many different uses across many applications) it's

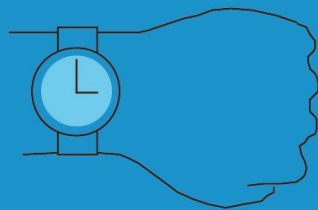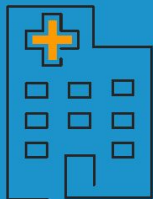https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#5b2ecb5d3aa9

Iaac Institute for advanced architecture of Catalonia

Buzzword and common perception. Is it or not?

more Open Data can help make
better decisions

2,549 hours
wasted
finding parking

7,000 lives
saved due to
quicker response

629 million
hours saved is
equivalent to
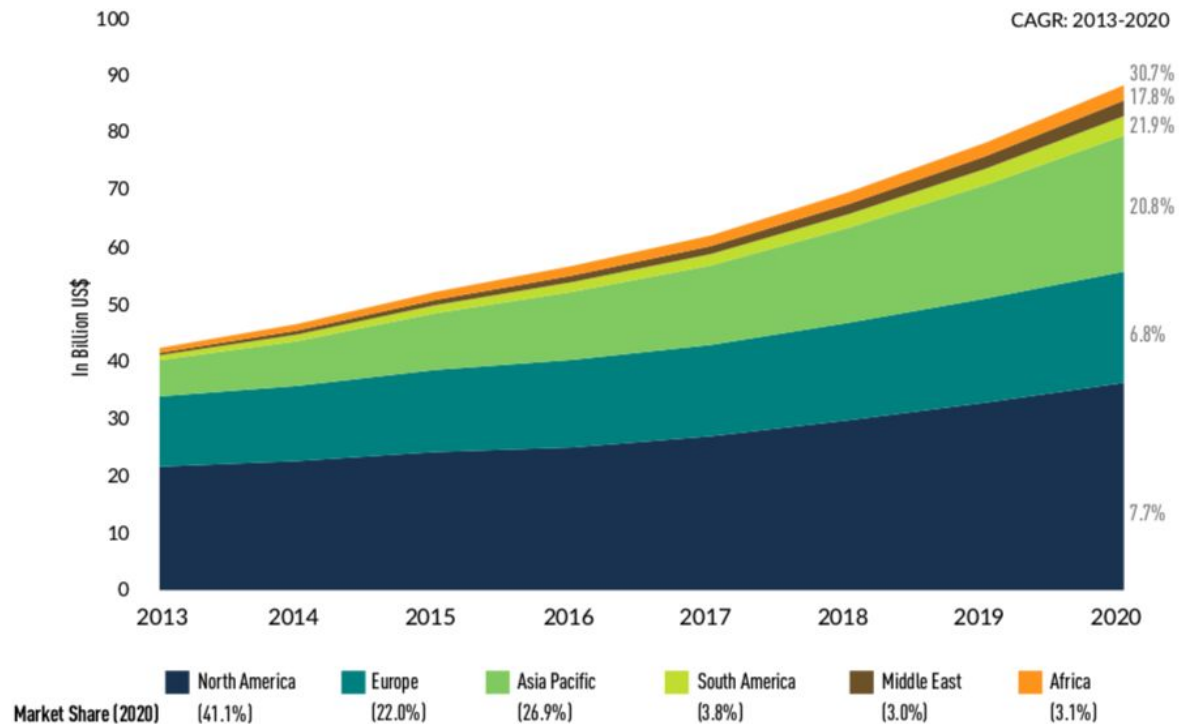**€ 27.9 bn**

5.5% less
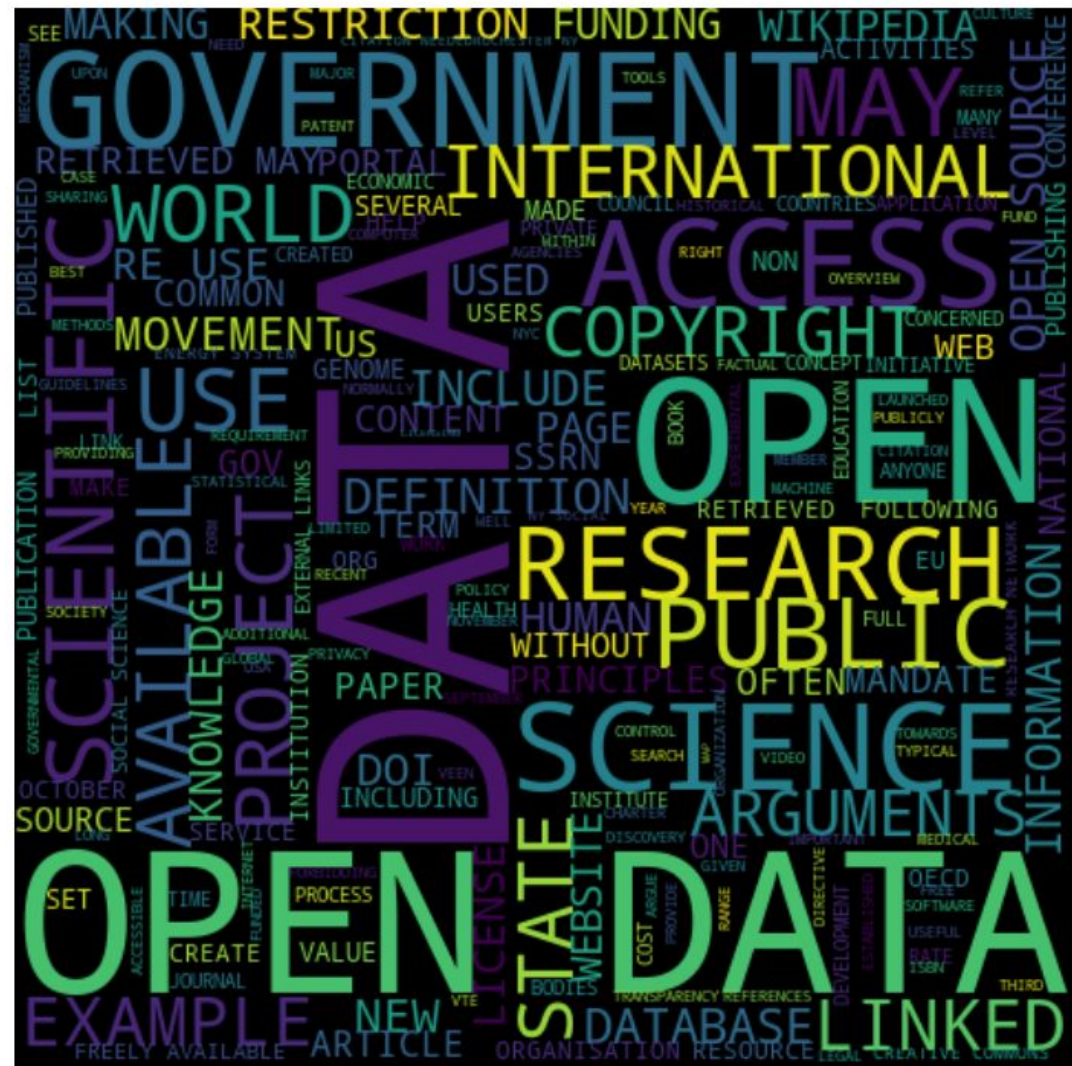road fatalities

Congestion
costs are
1% of GDP

16% less
less energy used

Do we need private data?

Text describing open data

Word cloud

Python library

https://python-graph-gallery.com/260-basic-wordcloud/

Iaac Institute for advanced architecture of Catalonia   **Do it yourself. Get some data and make it appealing**

Global

Global open data index index.okfn.org/
Natural earth www.naturalearthdata.com/
UN data (32 datamarts) data.un.org/
OECD (6k) data.oecd.org/
Global migration data analysis centre gmdac.iom.int/

Regional

EU portal (939k) www.europeandataportal.eu
Africa Geoportal www.africageoportal.com/
Inspire inspire-geoportal.ec.europa.eu/

National

USA (236k) www.data.gov/
Spain https://datos.gob.es/ www.idee.es/
UK (52k) data.gov.uk/
Ecuador (128) catalogo.datosabiertos.gob.ec

Local

Barcelona (452) opendata-ajuntament.barcelona.cat   www.bcn.cat/geoportal
NY opendata.cityofnewyork.us/
Amsterdam data.amsterdam.nl/
Medellin www.medellin.gov.co/geomedellin/

Open Data Portals, a pretty small selection of them

Wikipedia / DBpedia wiki.dbpedia.org/
Wikidata www.wikidata.org/
OSM wiki.openstreetmap.org/
Geonames https://www.geonames.org/
Zenodo (40k) zenodo.org/

Uber movement.uber.com
Here student sample data
Waymo waymo.com/open
City Bike www.citibikenyc.com/system-data

Technical

Connection, software, hardware, personal skills, general knowledge

Legal / Licensing

License agreements, terms of use, national security or privacy restrictions.

Pricing

From free to extremely expensive

Availability

In a readable format, means and access point, offered by producer or custodian

Scale

Spatial resolution (x,y), temporal (hms), value domains and coverage

Time

Covering the time period needed for the analysis

Google places developers.google.com/places

Twitter developer.twitter.com/

Yelp www.yelp.com/developers/

Open weather map https://openweathermap.org/api

Open Cage https://opencagedata.com/

Tomtom https://www.tomtommaps.com/mapdata/

Bing Maps https://blogs.bing.com/maps

ESRI https://www.esri.com/

Census Europe https://ec.europa.eu/CensusHub2

Copernicus https://scihub.copernicus.eu/

US census data https://www.census.gov

Humanitarian OSM tasks.hotosm.org/

Wikipedia www.wikipedia.org/

Open Sense Map opensensemap.org/

Smart citizen kit smartcitizen.me/kits/

Fablabs fablabs.io/

Moskito Alert mosquitoalert.com

Big Data

What do we mean by data?

THIS IS DIEGO'S PREPARATION FOR THE PYTHON PROGRAMMING COURSE AT IAAC;

Punch card

Created using the virtual keypunch: https://www.masswerk.at/keypunch/

Digital	See it *List & Plot*

	`my_df.head()`
	
	`my_df.plot()`

Read it
Pd.read
 	
`pd.read_csv('path/to/file.csv')`
	Query it
	a>3
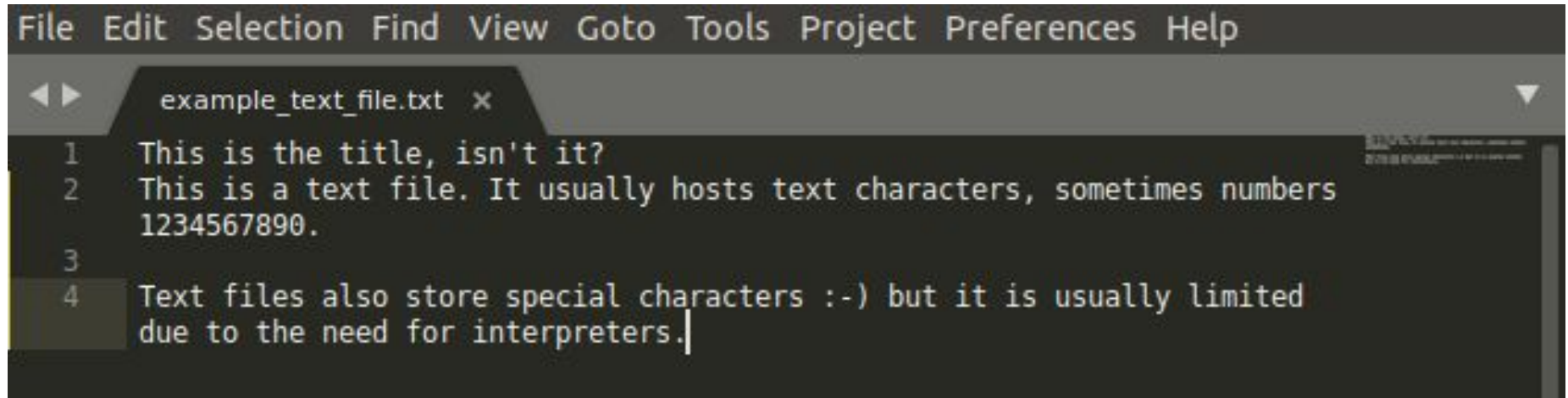	
	`if a > 4:`
	`    print('A is bigger than 3')`
	`else:`
	`    Print('No it is not')`

Open data defined right after the internet. In terms of Tim Berner Lee

https://5stardata.info/en/

Can we really access it?
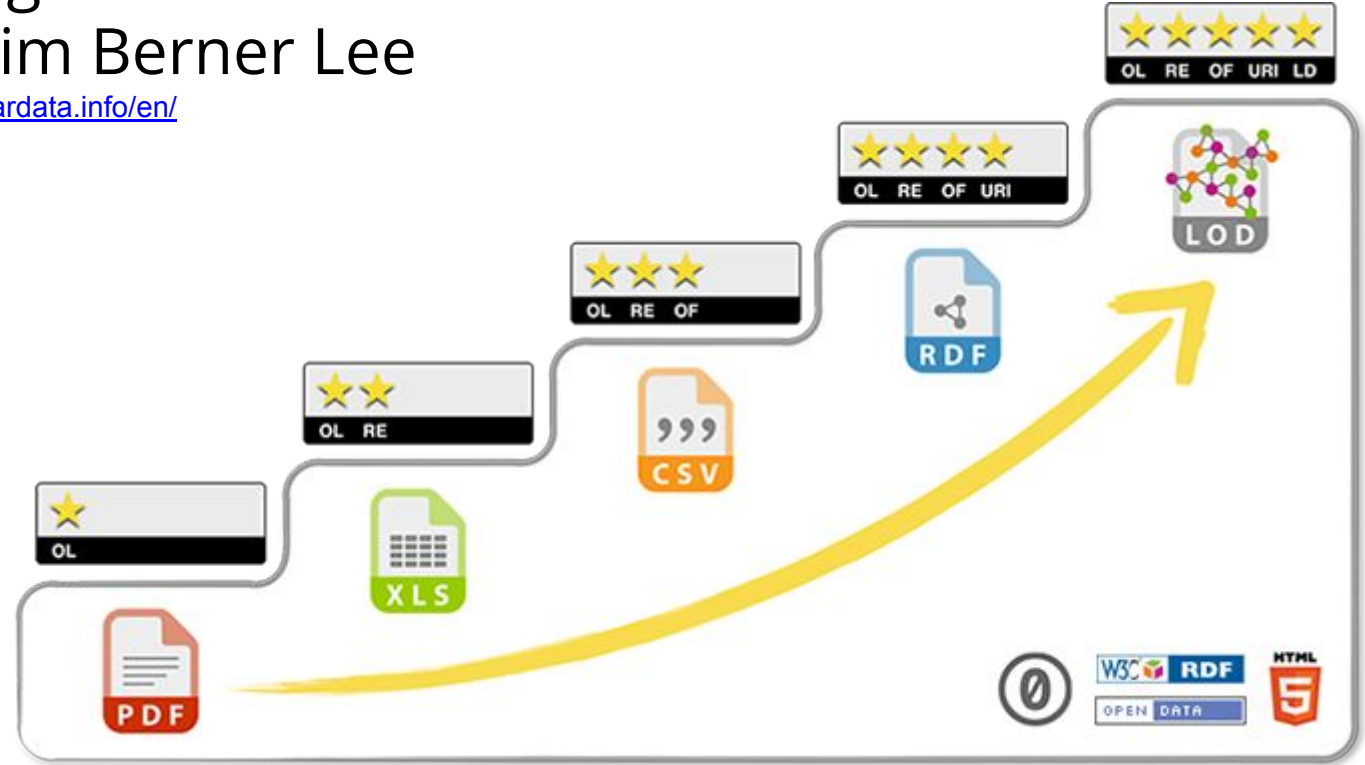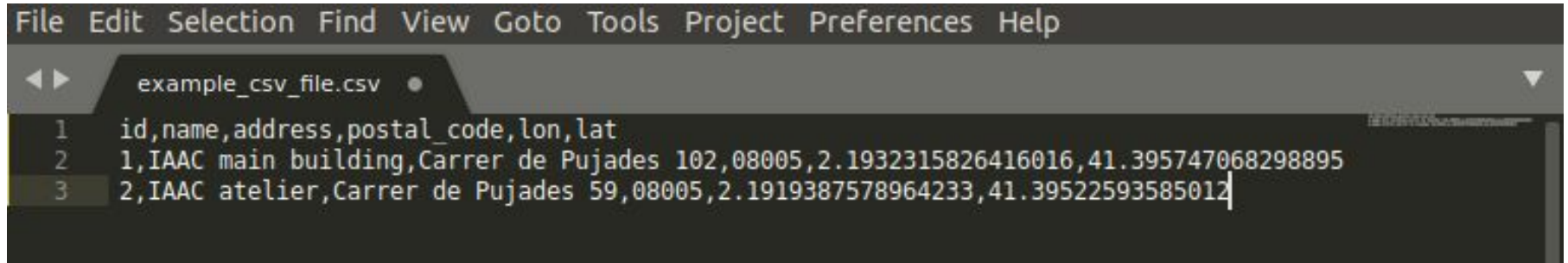
```
File  Edit  Selection  Find  View  Goto  Tools  Project  Preferences  Help

    ◄ ►    example_text_file.txt  ✕                                                    ▼

    1      This is the title, isn't it?
    2      This is a text file. It usually hosts text characters, sometimes numbers
           1234567890.
    3
    4      Text files also store special characters :-) but it is usually limited
           due to the need for interpreters.
```

Text Files   -   Extension .txt

You can store almost everything here... they are quite heavy.

Can you guess why?

File Edit Selection Find View Goto Tools Project Preferences Help

example_csv_file.csv

```
1   id,name,address,postal_code,lon,lat
2   1,IAAC main building,Carrer de Pujades 102,08005,2.1932315826416016,41.395747068298895
3   2,IAAC atelier,Carrer de Pujades 59,08005,2.1919387578964233,41.39522593585012
```

Comma separated Files   -   Extension .csv

This is a way to offer a more structured version of a text file.

It only works with simple structures, why?

Iaac Institute for advanced architecture of Catalonia

Microsoft Excel Files   -   Extension .xlsx .xls .ods

Open Structure based on grids, multiple data formats and functions. It is built on top rows, columns and sheets.

Data formats - XLSX Files

```json
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "properties": {
        "name": "IAAC main building",
        "address": "Carrer de Pujades 102",
        "postal_code": 08005
      },
      "geometry": {
        "type": "Point",
        "coordinates": [
          2.1932315826416016,
          41.395747068298895
        ]
      }
    },
    {
      "type": "Feature",
      "properties": {
        "name": "IAAC atelier",
        "address": "Carrer de Pujades 59",
        "postal_code": 08005
      },
      "geometry": {
        "type": "Point",
        "coordinates": [
          2.1919387578964233,
          41.39522593585012
        ]
      }
    }
  ]
}
```

Javascript object notation Files
Extension .json

It is a text file with a more complex structure optimised for data exchange through the web

Say HI to our friends "dictionaries" and "lists"?

Iaac Institute for advanced architecture of Catalonia **Data formats - JSON Files**

eXtended markup
language Files
Extension .xml / .{}ml

It is a text file with arrays
of markups designed for
machines to understand
web data

This is the weirdest so
far, isn't it?

Institute for
advanced
architecture
of Catalonia

We are hacking!!

Big Data

Data compilation

Go back to the compilation of data sources

Identify two or more cities/places in which you think an interesting problem deserves to be analysed

Download some data files and try to open it with the software you have in your computer

Create two or more repositories for cities

Barcelona, New York and others

What questions can we answer this data sets?

Can we create plots, maps, or other geospatial visualisation

Share some examples with your peers.

Iaac | Institute for advanced architecture of Catalonia | BARCELONA

MASTER IN CITY & TECHNOLOGY
DIGITAL TOOLS AND BIG DATA
2021/2022

FACULTY DIEGO PAJARITO