



Open data

Intelligent Infrastructure Design for the Internet of Things

Antonio Navarro

Index

- References
- Introduction
- Initiatives
- Open data formats
- Access to open data
- Conclusions

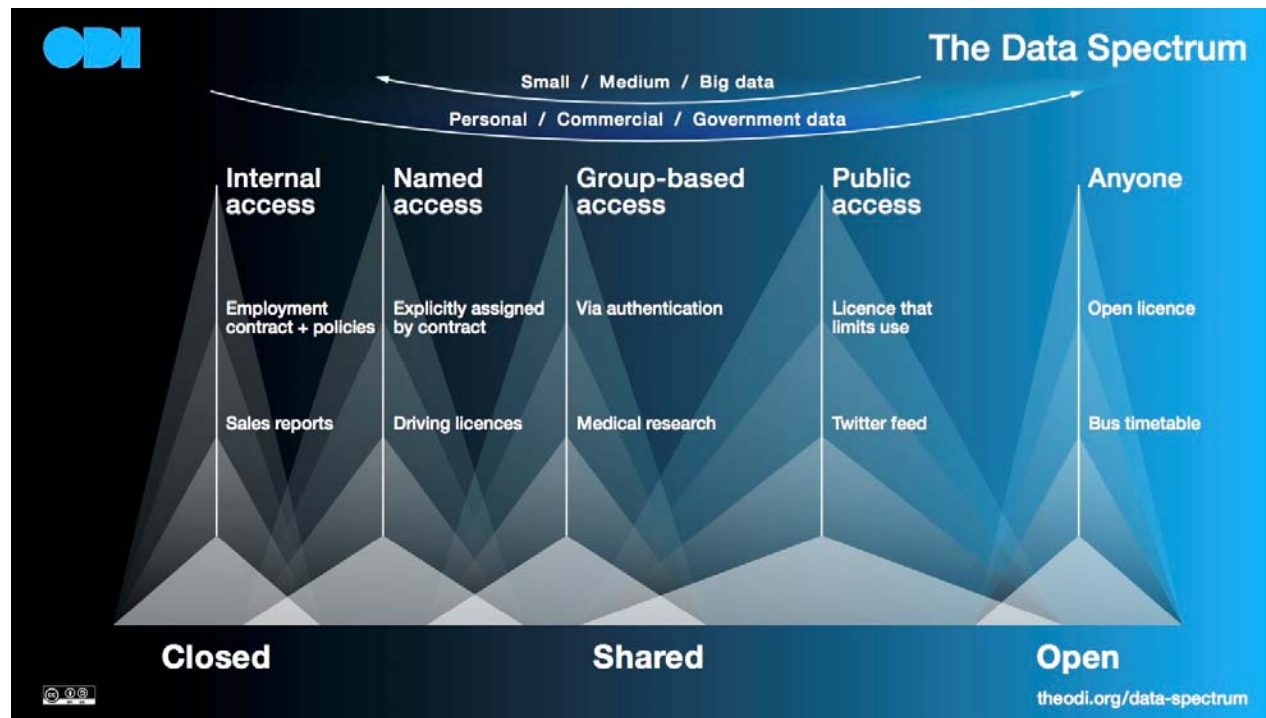
References

- (Hand, 2012) Hand, D. Open data is a force for good, but not without risks. *The Guardian*. <https://www.theguardian.com/society/2012/jul/10/open-data-force-for-good-risks>
- (Newcombe, 2015) Newcombe, T. The (Hidden) Cost of Open Data. *Governing*. <http://www.governing.com/columns/tech-talk/gov-open-data-cost-problems.html>
- (Roberts, 2012) Robert, T. The Problem with Open Data. *Computer Weekly*. <http://www.computerweekly.com/opinion/The-problem-with-Open-Data>

Introduction

- Open data initiatives seek to make public as much data as possible that can be useful to people.
- The information is published as datasets.
- Why do it?
 - Transparency
 - Providing social and commercial value
 - Participation and commitment
- The initiatives themselves consider that not all data can be public:

Introduction



Range of data according to the Open Data Institute

Initiatives

- There are basically two types of initiatives:
 - Those that promote open data:
 - Open Data for Development: <http://od4d.net/>
 - Open Data Institute (ODI): <https://theodi.org/>
 - Open Data Incubator Europe (ODINE): <https://opendataincubator.eu/>
 - Open Knowledge International: <https://okfn.org/>
 - The Opportunity Project: <https://opportunity.census.gov/>
 - Those that provide open data (in a way, these could be considered subtypes of the former):
 - Government of Spain: <http://datos.gob.es/>
 - US Government: <https://www.data.gov/>
 - GB Government: <https://data.gov.uk/>
 - Government of Russia: <http://data.gov.ru/>

Initiatives

- Government of Japan: <http://www.data.go.jp/>
- Autonomous Community of the Basque Country: <http://opendata.euskadi.eus/>
- Generalitat de Catalunya: <http://dadesobertes.gencat.cat>
- Madrid City Council: <http://datos.madrid.es>
- Tiétar Town Hall: <http://www.tietar.es/open-data>
- Governments that do not have open data initiatives:
 - China: <https://opendatachina.com/> (did not have in 2017).
 - North Korea
 - Cuba
 - Israel: <https://data.gov.il/> (did not have in 2017).
 - Turkey
 - Venezuela
- We have a list at: <https://www.data.gov/open-gov/>

Initiatives

- What can we find in the Spanish Government's open data portal?
- Let's take a look: <http://datos.gob.es>
- The scorecard is worth a look: <https://datos.gob.es/es/dashboard>
- The portal allows three different accesses:
 - Navigating through the portal itself
 - Through a REST API
 - Through a SPARQL point

Initiatives

- Are such initiatives worthwhile?
- It seems to be so, at least that's what the promoters say.
 - <https://www.england.nhs.uk/ourwork/tsd/data-info/open-data/examples/>
 - <https://theodi.org/projects-services/projects/>
- However, they have obvious problems (Hand, 2012; Robers, 2012):
 - Raw data can only be understood by a technical *elite*. This may cause more inequalities than benefits
 - To what extent are they *honest* initiatives or a matter of politicians' public relations?
 - Negative data (erroneous or not) can stigmatize groups/areas.
 - High collection costs

Initiatives

- The Government of Spain's own open data portal identifies five problems:
 - Licenses not open or unclear, which makes their use difficult
 - Incomplete data sets
 - Non-machine-readable or difficult to reuse formats
 - Outdated data
 - Access fees
- There is a major problem: the structure of the data, beyond the format.

Open data formats

- Open Knowledge International defines three key rules for making data open:
 - Keep it simple
 - Involve early and involve often.
 - Addressing common fears and misunderstandings
- It also identifies four key steps
 - Choosing data sets
 - Apply an open license
 - Making data available
 - Make them easy to find

Open data formats

- A key step is the publication of open data in a format easily processed by computers:
 - XML (eXtensible Markup Language)

```
< menu id="file" value="File">
  < popup>
    < menuitem value="New" onclick="CreateNewDoc()" />
    < menuitem value="Open" onclick="OpenDoc()" />
    < menuitem value="Close" onclick="CloseDoc()" />
  </popup>
</menu>
```

Open data formats

- JSON (JavaScript Object Notation)

```
{ "menu": {  
  "id": "file",  
  "value": "File",  
  "popup": {  
    "menuitem": [  
      { "value": "New", "onclick": "CreateNewDoc()" },  
      { "value": "Open", "onclick": "OpenDoc()" },  
      { "value": "Close", "onclick": "CloseDoc()" }  
    ]  
  }  
}}
```

Open data formats

- RDF (Resource Description Framework)

```
<? xml version="1.0" encoding="UTF-8"?>
  < rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:region="http://www.country-regions.fake/">
  < rdf:Description
rdf:about="http://en.wikipedia.org/wiki/Oxford">
    < dc:title>Oxford</dc:title>
      < dc:coverage>Oxfordshire</dc:coverage>.
      < dc:publisher>Wikipedia</dc:publisher>
    <region:population>10000</region:population>
    < region:principaltown rdf:resource="http://www.country-
regions.fake/oxford"/>
  </rdf:Description>
</rdf:RDF>
```

Open data formats

- Comma-Separated Values (CSV)
- Spreadsheet formats
- ESRI Shapefile (SHP): geographic information data
- Keyhole Markup Language (KML): geographic information data
- PC-Axis: statistical data
- PDF
- Text
- Scanned images

Access to open data

- Data is usually accessible in three different ways:
 - Navigating directly on the web
 - By means of web services, usually REST (Representational State Transfer).
 - POST: Create
 - GET: Read (e.g. <http://datos.gob.es/apidata/nti/territory/Province/Madrid>)
 - PUT: Update
 - DELETE: Delete
 - SPARQL (SPARQL Protocol and RDF Query Language)
 - It is a language for querying databases containing RDF data.
 - The DBPedia, for example, also has a SPARQL point: <http://dbpedia.org/sparql>.

Access to open data

- We can execute the following SPARQL query:

```
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dbp: <http://dbpedia.org/ontology/>
SELECT ? musician ? musicianName ? dateBirth ? dateDeceased
WHERE {
    ? musico dcterms:subject
    <http://dbpedia.org/resource/Category:Spanish_musicians>;
    rdfs:label ? nameMusician ;
    dbp:birthDate ? dateBirth ;
    dbp:deathDate ? deathDate .
    FILTER (LANG(? musicianName) = "en")
}
```

Conclusions

- Open data seems like a good idea
- However, it is necessary to distinguish useful data from mere political *posturing*.
- From an IT point of view, they do not represent a technical challenge, beyond choosing a suitable structure and format.
- Probably, the REST access mechanism is the most reasonable one.