Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing

Invited tutorial at ICTERI 2016
Kiev, 24/06/2016

Oscar Corcho
ocorcho@fi.upm.es
@ocorcho, @opencitydata_es
https://www.slideshare.com/ocorcho

License

- This work is licensed under the license CC BY-NC-SA 4.0 International
  - http://purl.org/NET/rdflicense/cc-by-nc-sa4.0
- You are free:
  - to Share — to copy, distribute and transmit the work
  - to Remix — to adapt the work
- Under the following conditions
  - Attribution — You must attribute the work by inserting
    - ”[source Oscar Corcho]” at the footer of each reused slide
    - a credits slide stating: ”These slides are partially based on “Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing” by O. Corcho”
  - Non-commercial
  - Share-Alike
Acknowledgements

This research has been funded by Ministerio de Economía y Competitividad (Spain) under the project "4V: Volumen, Velocidad, Variedad y Validez en la Gestión Innovadora de Datos" (TIN2013-46238-C4-2-R).

Tutorial Objectives

• Share some of our group experiences and recommendations in the application of Open Data principles for smart cities
  • Technical challenges and opportunities
  • Barriers for adoption
  • Required organisational changes for successful Open Data strategies

• With a special focus on how agreed vocabularies and shared data structures can make a difference
  • And a small bias towards how we are doing it in Spain
Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing

Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing
What is Open Data?

- Open data is data that can be *freely used, re-used and redistributed* by anyone - subject only, at most, to the *requirement to attribute and share alike*.

- Key aspects:
  - **Availability and access**: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the Internet. The data must also be available in a convenient and modifiable form.
  - **Re-use and redistribution**: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
  - **Universal participation**: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups.

Relevant Legislation. Europe and Spain

- **Open Access Initiative (2001)**. Scientific information; > 510 orgs
- **Aarhus Convention (1998)**. Right to participate and access; 41 countries and the EU
- **PSI Directives**. PSI reuse (2003/98/EC and 2013/37/UE)
- **Convention about access to official documentation (2009)**
  - 12 countries

- **Law 37/2007**. PSI reuse (transposition of directive 2003/98/EC)
  - Modified in law 18/2015 (BOE 10/07/2015, directive 2013/37/UE)
- **Law 11/2007**. Citizen access to public services, and rights to good quality services
- **RD 4/2010** Esquema Nacional de Interoperabilidad
  - Open standards, technology neutral, open source
- **RD 1495/2011** Desarrolla la Ley 37/2007, para el ámbito estatal
- **Norma Técnica de Interoperabilidad** (19/02/2013, BOE 4/3/2013)

An Explosion of Open Data Portals

Open Data and how to publish it

1) In a posterboard
   • For those with a lot of free time available
   • Or those who happen to be there at the right time

Adapted from Antonio Rodríguez Pascual (IGN)
Open Data and how to publish it

2) On a Web page or mobile app
   - For people

Adapted from: Antonio Rodríguez-Pascual (IGN)

Open Data and how to publish it

2) On a Web page or mobile app
   - But generated from structured formats (e.g., GTFS)

Adapted from: Antonio Rodríguez-Pascual (IGN)
3) In files

- These can be loaded by humans in information systems (XML, HTML, CSV, GTFS, etc.)
- Luckily, it is not a scanned PDF

4) Via Web Services

- They can be used by systems (and sometimes persons)
- They allow generating added value
- Ease of integration in the application logic
All together..., Shaken, not stirred...

Open Data and how to publish it

5) Via APIs (semantically enhanced) and linked
• To be used by systems (and sometimes persons)
• It allows generating added-value services
• Standardised formats (JSON, JSON-LD, RDF)
• Standardised models (vocabularys, ontologies)
And many more: JSON, JSON-LD, Shapefiles, KMZ, KML, PC-Axis, etc.

Recap: The 5-star categorisation from Tim Berners-Lee

- Make your stuff available on the Web (whatever format) under an open license
- Make it available as structured data (e.g., Excel instead of image scan of a table)
- Use non-proprietary formats (e.g., CSV instead of Excel)
- Use URIs to identify things, so that people can point at your stuff
- Link your data to other data to provide context
Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing
Why my interest in cities

- Regional and national datasets are useful to provide indicators, comparisons across countries, etc.
- However, local datasets are usually better for the generation of economical and/or social value
  - Closer to people
  - Closer to businesses
  - Closer to developers

The current status of Open City Data in Spain

Source: CTIC and OKFN
Open Data Portals in Spain

• Around 60 open data portals for Spanish cities (and growing)
  • Gijón, Santander, Bilbao, Azkoitia, Azpeitia, San Sebastián, Pamplona, Zaragoza, Lleida, Manresa, Barcelona, Badalona, El Prat, Cornellá, Viladecans, Gavá (and a few more in Barcelona’s province), Tarragona, Valladolid, Alcobendas, Madrid, Cáceres, Valencia, Lorca, Málaga, Arona, etc.
  • Also regional open data portals
    • Many of which aggregate local data
    • Cabildo de Tenerife, Diputación de Málaga, Gobierno de Aragón, etc.
  • And national institutions, public companies, etc.

Relevant legislation for cities

• City Ordinances on Transparency, Access and Reuse of Information
  • Zaragoza: [http://www.zaragoza.es/ciudadania/gobierno-abierto/participar/detalle_Normativa?id=3983](http://www.zaragoza.es/ciudadania/gobierno-abierto/participar/detalle_Normativa?id=3983)
    • Initial draft (25/06/2012)
    • Final approval (01/04/2014)
  • Template from the Spanish Federation of Municipalities and Provinces - FEMP (may 2014)
    • [http://www.femp.es/files/11-5133-fichero/Ordenanza%20Transparencia,%20Acceso%20y%20Reutilizaci%C3%B3n%20Informaci%C3%B3n.pdf](http://www.femp.es/files/11-5133-fichero/Ordenanza%20Transparencia,%20Acceso%20y%20Reutilizaci%C3%B3n%20Informaci%C3%B3n.pdf)
  • 2015. UNE 178301. Ciudades Inteligentes. Datos abiertos (Open Data). MINETUR Y AENOR
    • More about this to come later in this presentation
Technology that is used

- Ad-hoc developments over CMSs
  - Zaragoza (http://datos.zaragoza.es/)
  - Madrid (http://datos.madrid.es/)
  - Gijón (http://datos.gijon.es/)
  - http://datos.gob.es
- CKAN
  - The most recent ones, in general (e.g.,
- Socrata
  - Rubí, Gavà, etc.
- More details (in Spanish) at

Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing
Building Agreed Vocabularies and Data Structures for **Successful** Open City Data Sharing

### Main challenges for success

- Many efforts so far focused on setting up the technology infrastructure
  - That's ok, but content is probably more important
- Heterogeneity on the selection of datasets, their formats and granularity
  - Build an app, deploy everywhere
- Open data portals are at the end of the data production chain
  - They are not an integral part of data management inside cities

*Note: the following slides are based on my Semantics 2015 dinner speech on “Slow-cooked data and APIs”*
Rule 1.

Chop your onions appropriately

- Take care about the number of datasets that you produce
  - There’s still a silly competition about “my open data portal has more datasets than yours”
    - This provokes, sometimes, over-segmentation of data
- Main question: What makes a dataset useful and which datasets should I publish?
Rule 1. Chop your onions appropriately

- UNE 178301:2015
  - Norm on Open Data for Smart Cities
- Organised by
  - AENOR CTN 178 group
    - Government and Mobility
      - Government
      - Open Data (led by Localidata)
- Formed by
  - Several cities
  - Private companies
  - Nation-wide organisations

10 datasets selected
- Based on frequency of requests from reusers

Slow but steady adoption for 2015

And now working on extending it to 100 datasets
- With an additional group of people

Datasets

<table>
<thead>
<tr>
<th>Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultural Agenda</td>
</tr>
<tr>
<td>Traffic</td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Streets</td>
</tr>
<tr>
<td>Public Transport</td>
</tr>
<tr>
<td>Touristic Places and POIs</td>
</tr>
<tr>
<td>Budget</td>
</tr>
<tr>
<td>Shop Census</td>
</tr>
<tr>
<td>Air Quality</td>
</tr>
<tr>
<td>Contracts</td>
</tr>
<tr>
<td>Parkings</td>
</tr>
</tbody>
</table>
Rule 2. Add some spices, but not too many

- Annotate (semantically) your data, so that others can understand what you produce
  - And produce examples for consumers to understand them

- Don’t wait until all schema.org properties are settled
- Generate SKOS thesauri for your own classifications
  - e.g., for groups of citizens (young, elderly, etc.), for types of events (cultural, children, music, etc.)
Rule 3. Try different ways of plating up your food

- Produce your data in different formats
  - Agreed-upon JSONs
  - JSON-LD
  - RDF
  - Agreed-upon CSVs
    - With the upcoming CSV on the Web

- But don’t get crazy at offering all options
  - The ones that get finally used are more than enough
Rule 4.
Let children appreciate (and cook) open data

Let children understand the benefits of open data (and Citizen Science)
and how they can contribute to improving the data of their city.
Rule 5…
Eat your own dog food

Eating Your Own Dog Food

Let’s better say…

Rule 5.
Try it out yourself first…

… Before giving your food to your customers
Rule 5. Try it out yourself first…

- Open data by default
- So that your applications are also based on open data

Source: Los Datos Abiertos como Eje Central del desarrollo de la Plataforma de Gobierno Abierto. M.J. Fernández-Ruiz, V. Morlán

Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing
Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing

Whom of you has never ever eaten a burger in his/her life? Well, if you are vegetarian, we can think of something else.

Rule 6.
Fast food has its value as well, why not...

You go anywhere in the world and know how McDonald’s burgers are...

So let’s only learn this from fast food..
Rule 6. Fast food has its value as well, why not...

- When we open our data, let’s use at least the same data structures

I want to publish my data

- I am using GTFS
- I am using my own CSV structure
- I provide it as a Web service

Write an app and deploy everywhere
**UNE 178301:2015. Context**

- UNE 178301:2015
  - Norm on Open Data for Smart Cities
- Organised by
  - AENOR
- AENOR CTN 178 group
  - Government and Mobility
    - Government
    - Open Data (led by Localidata)
- Formed by
  - Several cities
  - Private companies
  - Nation-wide organisations

**UNE 178301:2015. Objectives**

- Establish **metrics and indicators** that allow evaluating the **degree of maturity** of open data elaborated or curated by the public sector…
- … so that we can **facilitate reuse**,
- …
- … and define a minimal list of **datasets** that should be published by all open city data initiatives, together with their **associated vocabularies**
UNE 178301:2015. The team

- Coordinator
  - Esther Minguela (Localidata)
  - Oscar Corcho (OEG-UPM and Localidata)
- 35 members who belong to...
  - Medium&Large Cities (10) – mostly City Information Managers
  - Private companies working for the public sector (6)
  - Regions (3) – mostly Region Information Managers
  - Ministries or alike (3)
  - Geographic sector (3)

- The core of the work done over a period of 6 months

UNE 178301:2015. The team

- Coordinators
  - Localidata. Esther Minguela y Oscar Corcho
- Participants
  1. Animsa. Iñigo Sancho
  2. Animsa. Jesús Vera
  3. Ayuntamiento de Albacete. Manuel Tobarra
  4. Ayuntamiento de Bilbao. Teresa Alba
  5. Ayuntamiento de Bilbao. Josu Santacruz
  6. Ayuntamiento de Burgos. José María Díez
  7. Ayuntamiento de Madrid. Enrique Crespo
  8. Ayuntamiento de Madrid. José Luis Cano
  9. Ayuntamiento de Rivas. Jesús García
  10. Ayuntamiento de Segovia. Alberto Gómez
  11. Ayuntamiento de Zaragoza. María Jesús Fernández
  12. Comunidad de Madrid. Inmaculada Sánchez
  13. Comunidad de Madrid. Concepción García
  14. IGN. Antonio Rodríguez
  15. IGN. Celia Sevilla
  16. COIT. Jorge Díaz
  17. CRTM. Concha Chapa
  18. CRTM. José Antonio Cascales
  19. Etra. Vicente Sebastián
  20. FEMP. Pablo Bárdenas
  22. Gmv. Antonio Velasco
  23. Indra. Elisabet Terrades
  24. Indra. Jordi Marín
  25. Junta de Castilla y León. Antonio Ibáñez
  26. MINETUR. Ana Pérez
  27. MINETUR. Javier García
  28. MINHAP. Aleida Alcaide
  29. Proconsi. Constantino Lázaro
  30. Red.es. José Ignacio Sánchez
  31. Red.es. Sonia Castro
  32. SEGITTUR. Calixto Mellen
  33. SEGITTUR. Enrique Lanzos
  34. SEGITTUR. Rosa Muñoz
  35. SEGITTUR. Susana García
  36. Universidad de Zaragoza. Javier López
UNE 178301:2015. Important dates

- **May 2013.** Proposal to work on Open Data is sent to the AENOR CTN 178 group
- **December 2013.** Approval to start working on it
- **June 2014.** Norm available for public comments
- **September-October 2014.** Public comments reception
  - Around 100 comments received
- **November 2014.** Proposed for approval to committee CTN178
  - Wide consensus
- **January 2015.** Publication as UNE 178301:2015
- **2016.** Norm being proposed by the Spanish SETSI to ITU Study Group 20 on “Internet of Things for Smart Sustainable Cities”

UNE178301:2015. Metrics

[Diagram showing domains and dimensions with metrics]
37 Metrics, grouped in domains (and dimensions)

**Strategic Domain**
1. Strategy
2. Leadership
3. Service-level agreement
4. Sustainability

**Legal Domain**
5. External and internal legal norms
6. Usage and licensing conditions

**Organisational Domain**
7. Responsible unit
8. Skilled team
9. Inventory of data
10. Priority
11. Measurement of the process
12. Measurements of usage and impact

**Technical Domain**
13. Catalogue
14. Available in the public sector catalogue
15. Documented datasets
16. Categories and search facilities
17. Availability
18. Persistent and friendly references
19. Accessibility
20. Access for free
21. Access systems in place
22. Primary data
23. Completeness
24. Documentation of data
25. Correctness
26. Geo-referencing
27. Linked Data
28. Update processes
29. Update frequency
30. New dataset inclusion
31. Data quantity
32. Data format
33. Vocabularies

**Economic and social domain**
34. Transparency, participation and collaboration
35. Complaint/Conflict management
36. Fostering reuse
37. Developed reuse initiatives

---

**UNE 178301:2015. Scoring**

1. **Scoring**

<table>
<thead>
<tr>
<th>Level</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 0. Non existing</td>
<td>0</td>
</tr>
<tr>
<td>Level 1. Starting</td>
<td>1</td>
</tr>
<tr>
<td>Level 2. Consolidated</td>
<td>2</td>
</tr>
<tr>
<td>Level 3. Advanced</td>
<td>3</td>
</tr>
</tbody>
</table>

2. **Weight inside each domain**

<table>
<thead>
<tr>
<th>Strategic domain</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td>25 %</td>
</tr>
<tr>
<td>Leadership</td>
<td>50 %</td>
</tr>
<tr>
<td>Service level agreements</td>
<td>10 %</td>
</tr>
<tr>
<td>Sustainability</td>
<td>15 %</td>
</tr>
</tbody>
</table>

3. **Total value calculation**

\[ \text{Value} = ((\text{Score} \times \text{Weight})/3) \times 100 \]

4. **Open data maturity indicator**

<table>
<thead>
<tr>
<th>Total Value</th>
<th>0-200</th>
<th>201-400</th>
<th>401-600</th>
<th>601-800</th>
<th>801-1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Data Maturity Indicator</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
UNE178301:2015. High-Priority Datasets

- Listing based on the current inventories from all cities (and regions)
- Harmonisation
- Votes according to PSI-reuse requests

<table>
<thead>
<tr>
<th>Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultural Agenda</td>
</tr>
<tr>
<td>Traffic</td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Streets</td>
</tr>
<tr>
<td>Public Transport</td>
</tr>
<tr>
<td>Touristic Places and POIs</td>
</tr>
<tr>
<td>Budget</td>
</tr>
<tr>
<td>Shop Census</td>
</tr>
<tr>
<td>Air Quality</td>
</tr>
<tr>
<td>Contracts</td>
</tr>
<tr>
<td>Parkings</td>
</tr>
</tbody>
</table>

UNE 178301:2015. Vocabularies

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tourism</td>
<td>Vocabulary: <a href="http://vocab.linkeddata.es/datosabiertos/def/turismo/lugar">http://vocab.linkeddata.es/datosabiertos/def/turismo/lugar</a></td>
</tr>
<tr>
<td>Budget</td>
<td>Vocabulary: <a href="http://vocab.linkeddata.es/datosabiertos/def/hacienda/presupuesto">http://vocab.linkeddata.es/datosabiertos/def/hacienda/presupuesto</a></td>
</tr>
<tr>
<td>Public tendering</td>
<td>Vocabulary: <a href="http://contem.unizar.es/de/sector-publico/pproc">http://contem.unizar.es/de/sector-publico/pproc</a></td>
</tr>
<tr>
<td>Air quality</td>
<td>Vocabulary: <a href="http://www.w3.org/2005/Atom/atomProfiles">http://www.w3.org/2005/Atom/atomProfiles</a></td>
</tr>
<tr>
<td>Traffic</td>
<td>Vocabulary: <a href="http://vocab.linkeddata.es/datosabiertos/def/transporte/trafico">http://vocab.linkeddata.es/datosabiertos/def/transporte/trafico</a></td>
</tr>
<tr>
<td>Public transport</td>
<td>Vocabulary: <a href="http://vocab.linkeddata.es/datosabiertos/def/transporte/transportePublico">http://vocab.linkeddata.es/datosabiertos/def/transporte/transportePublico</a></td>
</tr>
<tr>
<td>Parking</td>
<td>Vocabulary: <a href="http://vocab.linkeddata.es/datosabiertos/defurbanismo-infraestructuras/aparcamiento">http://vocab.linkeddata.es/datosabiertos/defurbanismo-infraestructuras/aparcamiento</a></td>
</tr>
</tbody>
</table>
Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing

Where are we heading to now?

http://ojoaldata100.okfn.es/
Technology support for this activity

• Larger-scale vocabulary development
  • Hence, tools are needed to support this effort more systematically

• A repository for vocabularies
  • https://github.com/opencitydata

• Ontology development tools using GitHub
  • http://ontology.linkeddata.es/
  • https://github.com/mariapoveda/vocab.linkeddata.es
  • http://oops.linkeddata.es/
  • https://github.com/dgarijo/widoco

In summary...

• Open data offers good opportunities to cities
  • For 3rd party reusers
  • For city officers

• But these efforts need to be taken seriously in order to make the most of it

• We need a principled approach to publish and exploit open data more effectively
  • Not only based on open formats and 5-star classifications
  • But also on shared vocabularies and data structures
    • Develop once, deploy everywhere
    • Long live to ontology engineering
Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing

Invited tutorial at ICTERI 2016
Kiev, 24/06/2016

Oscar Corcho
ocorcho@ifi.upm.es
@ocorcho, @opencitydata_es
https://www.slideshare.com/ocorcho