Open Data in Practice

http://training.theodi.org/InPractice

David Tarrant · @davetaz
Introductions

Your name

What is your favourite example/use of open data?

What do you want to do differently after the course?
Build a solid foundation and experience in publishing, consuming and building a business in Open Data.
Schedule

Day 1: Practical publication

Day 2: Business, the law and open data

Day 3: Enriching and visualising data
Agenda - Today

The characteristics of data
Data discovery patterns
*** Lunch ***
Data publication platforms
Quick big data break
Practical publication hands-on
Recap session
Exercise

What is Data?
Value

- Data
- Information
- Knowledge
Exercise

What is Open Data?
Option A

Open data is data that is made available by organisations, businesses and individuals for anyone to access, use and share.

- Open Data Institute
  Introduced November 2014
|-----------------------------------------|----------------|---------------------------|-------------------------------------------|-----------|---------------|--------|---------------|----------------|--------------------|----------------|
Open data is hardly ever appropriately licensed.

Source: Thomas Levine
Warning: This result was actually based upon the wrong date field for updates. I personally think you’ll find a similar result with the right date field being used.

Source: Thomas Levine
Open data is rarely structured.

Source: Thomas Levine
Publication phases

**Phase 1:** Get the data online, in some form. This will help with the trust and transparency and community building.

**Phase 2:** Increase the usability of the data by potentially publishing differently and keeping it up to date.
Today’s mission

To move to phase 2 of publishing open data and solve some of the phase 1 problems.

What best practice guidelines and tools will help us achieve this phase 2 goal?
5-Stars

http://5stardata.info/
Open Data Certificate

http://certificates.theodi.org
Introducing Open Refine

http://openrefine.org
Session 1

The characteristics of data
Identify a number of different characteristics of data

Explain the justifications for publishing different types of data

Evaluate the current open data ecosystem and future opportunities
Exercise (part 1)

In your pre-training exercise, you were all asked to identify a dataset.

In your groups briefly discuss each others datasets and write down some key characteristics of each.

Also write the dataset title on a post-it, one per post-it.
Types of Data

Reference data

“things”

Transaction data

“stats involving things”
Exercise

Categorize your data into reference and transactional data.

If they are all in one category you have 2 minutes to add some new datasets to the empty category.

When done, put a “T” or and “R” on each dataset post-it.
Types of Data

Reference data
  "things"

People  Facilities  Places
Books  Buildings

Transaction data
  "stats involving things"

Expenditure
Weather Consumption
Observation

£

Climate

Car
Update frequency

Static  In frequent updates  Frequent updates  Live
Exercise

Categorize your data into **frequency of updates**

If they are all in one category you have 2 minutes to add some new datasets to the empty category/ies

Put a number on your post-its representing the frequency of updates.

0 = static, 1 = In frequent, 2 = Frequent, 3 = Live
Data Representations

Tabular

<table>
<thead>
<tr>
<th>Region</th>
<th>Production (thousand MT)</th>
<th>Change from last year</th>
<th>Change from 5 year average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>57,099</td>
<td>-0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Mato Grosso</td>
<td>10,096</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Parana</td>
<td>9,571</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Sao Paulo</td>
<td>9,844</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Goias</td>
<td>6,859</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Mato Grosso do Sul</td>
<td>4,296</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Mato Grosso</td>
<td>2,262</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Minas Gerais</td>
<td>2,543</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Sao Paulo</td>
<td>1,382</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Maranhao</td>
<td>1,090</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Santa Catarina</td>
<td>9,639</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Roraima</td>
<td>987</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Para</td>
<td>966</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Distrito Federal</td>
<td>106</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Roraima</td>
<td>966</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Hierarchical

Network/Graph
Exercise

Categorize your data into tabular, hierarchical (tree) and graph (network)

If they are all in one category you have 2 minutes to add some new datasets to the empty category.

Add the word “tab”, “tree” or “net” to your post-its to represent the different structures.
Trust and Transparency | Enabling the economy
One more

Categorize your data into transparent and enabling.
Summing up

Do you have any obvious grouping of your datasets?

Is this reflective of the whole open data ecosystem?
G8 Open Data Charter and Technical Annex

Published 18 June 2013

Contents

1. Principle 1: Open Data by Default
2. Principle 2: Quality and Quantity
3. Principle 3: Usable by All
4. Principle 4: Releasing Data for Improved Governance
5. Principle 5: Releasing Data for Innovation
6. Technical annex
Exercise

Pick one “group” of datasets that share similar colours and come up with a data publication strategy for getting these datasets online and usable.

What are the publication requirements on the human publisher?
What are the requirements on potential users?
Outcomes

Identify a number of different characteristics of data

Explain the justifications for publishing different types of data

Evaluate the current open data ecosystem and future opportunities
Thank-you