Assessing Accessibility

Usability & Infrastructure of Open Data Portals

Umberto Boschi, Tommaso Elli, Phillip Morris, Oriane Piquer-Louis, Guillén Torres, Jeroen de Vos
Intro

The Open Data you need lives HERE
Portals mediate access to data

They have their own particular discourse

Each portal makes a ‘promise’ to the user
Research Question

How can we measure the accessibility of OpenData Portals?
(with digital methods)

● “The frontend”
  ○ Semantic analysis: what are portals focused on?
  ○ Spatial analysis: relative relevance of the different sections in the portals
  ○ Code intricateness: is it correlated with visual intricateness?
  ○ Usability.gov-style analysis.

● Online talk?
  ○ Is the Open Data community voicing satisfaction/dissatisfaction online?

● Infrastructure
  ○ Programming Issues: how messy is the Portal’s infrastructure?
Seven portals analyzed (frontend and online talk)

1. Data.gov.uk (United Kingdom)
2. Data.gouv.fr (France)
3. Data.gov.au (Australia)
4. Dati.gov.it (Italy)
5. Datos.gob.mx (Mexico)
6. Data.gov (USA)
7. Opencanada.org (Canada)

Plus one case study about infrastructure issues → gov.uk (best rated according to openberometer.org + full source code on Github)
Research question: what do the portals promise to their readers?

1. Quick semiotic analysis of the frontpage to see what is the first impression on the reader = what the website actually means?
2. Scrape all words on the front page
3. Co-occurrence analysis (with community detection) with Iramuteq

**Result:** discourse/promise of the portal
Semantic analysis

datos.gob.mx (México)
data.gouv.fr (France)
Result: How do institutions organise content across their portal home pages?
Recognised 8 different “types” of section: header, search, guide, tools, news, socials, statistics, footer.
1. Search for specific keywords on the code
2. Plot screen-captures according to code intricateness.

**Result:** Does the structure of the code relates to visual intricateness?
Information Architecture analysis

Adapted Usability.gov analysis based on Organization Structure and Organizational Scheme:

1. Map the structure of links present on the front page
2. Analyze the link arrangement on the front page

Result: Does the amount, structure and presentation of links affect data portal usability?
Methods > “the frontend” (4/4)

Information Architecture - Organization Structure analysis

**Result:** UK https://data.gov.uk/ 45 total links 44 internal links, 1 external links
Information Architecture - Organizational Scheme analysis

**Result:** Navigating by the categories given on the front page will not include most datasets

<table>
<thead>
<tr>
<th>Nation</th>
<th>Open Data Portal</th>
<th>Total Links</th>
<th>Total Datasets</th>
<th># co2 Datasets by search</th>
<th>Total Categories</th>
<th>Icons</th>
<th>Total Datasets in Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td><a href="http://www.data.gouv.fr/">http://www.data.gouv.fr/</a></td>
<td>269</td>
<td>27,437</td>
<td>21</td>
<td></td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Italy</td>
<td><a href="http://www.dati.gov.it/">http://www.dati.gov.it/</a></td>
<td>127</td>
<td>18,222</td>
<td>59</td>
<td>13</td>
<td>Yes</td>
<td>6,675</td>
</tr>
<tr>
<td>Mexico</td>
<td><a href="https://datos.gob.mx">https://datos.gob.mx</a></td>
<td>N/A</td>
<td>25,804</td>
<td>626</td>
<td>11</td>
<td>Yes</td>
<td>59 (2,280)</td>
</tr>
<tr>
<td>United Kingdom</td>
<td><a href="https://data.gov.uk/">https://data.gov.uk/</a></td>
<td>45</td>
<td>41,346</td>
<td>28,421</td>
<td>12(10)</td>
<td>No</td>
<td>39,136</td>
</tr>
</tbody>
</table>
Result: The UK deemphasizes the categories of Crime & Justice and Defense
Online Talk and Network Analysis

How is the usage of the portals reflected on social networks (Twitter)?

Which kind of actors are discussing open data on social networks?

Can we find evidence of data portals usage by citizens? Are there any negative experiences? (spoiler: we couldn’t find any)
Online Talk and Network Analysis

Process:

- Collecting the data.
- Gephi: identifying clusters.
- Magnifying and dissecting the clusters: describing the nodes, using a sample cluster.
- Close reading: Isolating “negative” tweets and identifying dissenting voices.
This is a guy hijacking the hashtag with great success: @PleasureEthics
Actors:
- Institutions
- Non-Profits
- Media
- Banks
- Tech companies
- Open data municipality projects
What about the negative experiences?

- Selected from the corpus the tweets containing sad/angry emojis.
- Got 320 tweets against a total of 1.600.000
- A significant percentage of those are complaints about the (temporary) removal of the US data after the election of Trump.
- Most of them are people expressing frustration about working with data or attending open data events.
Issue analysis

1. Scraping the open/closed issues of government projects on Github related to Open Data (14 international projects, the whole gov.uk website)
2. Co-occurrence analysis (with community detection) with Iramuteq
3. Issue-Label analysis (frequency

Result: A sneak peek of what programmers discuss and consider relevant
Methods > “the backend”

Issue analysis: gov.uk

Subset of issues related to accessibility
Issue analysis: gov.uk

Issue labels: frequency (size) and number of comments (color)
Conclusion

● To get a picture of the accessibility of a portal, it is necessary to combine different types of methodologies (only one doesn’t make it)
● Although most portals look alike (international standards), their discourse and the type of relationship they establish with the user are very different.
● There seems to be a relation between how complicated the infrastructure of a portal is and its accessibility.
So long and thanks for all the fish!

THE #1 PROGRAMMER EXCUSE FOR LEGITIMATELY SLACKING OFF:

"DMI-TCAT IS BUSY!"

HEY! GET BACK TO WORK!

TCAT BUSY!

OH, CARRY ON.
Methods > “the frontend” (4/4)

Information Architecture - Organization Structure analysis

**Result:** UK https://data.gov.uk/ 45 total links 44 internal links, 1 external links
Information Architecture - Organization Structure analysis

**Result:** U.S.A.  https://www.data.gov/  77 links  56 internal links, 21 external links
Information Architecture - Organization Structure analysis