Data Innovation in Official Statistics: the Leading Role of Open Data

Giorgio Alleva
President
Italian National Institute of Statistics - Istat
“Open data are adding a new dimension to big data analytics and giving rise to novel, data-driven innovations.”

Official Statistics meets Open Data

- Official Statistics can more easily reach such wide range of users if conveyed through open data
- Recent technologies advances in the open data community enable **new advanced dissemination channels for Official Statistics**

**Benefits**

- Reinforce trust in NSO role
- Getting closer to users
- Reaching new users
- Give back information to users
- Information Enrichment (Metadata)
Linked Open Data

Semantic Web
Technological Standards

+ 

Knowledge Representation

= 

Linked Open Data - LOD
The Linked Open Data (LOD) portal allows to access and browse data in open format based on semantic web technologies and standards. The LOD portal can be queried directly from any application and responds to the need, expressed by the community of users, to have standardized and interoperable data.

Istat LOD Portal: [http://datiopen.istat.it](http://datiopen.istat.it)
Istat’s Linked Open Data Portal - 2

**Interaction**

- **Human basic**
  - Guided Access
  - Download
  - Guided queries

- **Human Advanc.**
  - Query via SPARQL EndPoint
  - Navigation

- **MachineTo Machine**
  - Web Service
  - Query REST on SPARQL EndPoint

---

**Access freedom**

UN World Data Forum, Cape Town, 15 -18 January 2017
Use Case 1: Spatial Querying

LOD when accompanied by spatial information allow to access data using spatial queries

App that displays on a map some population indicators of the nearest census sections to specific GPS coordinates
Use Case 2: Federated Querying - 1

With LOD, it is very easy to realize analyses with comparison of data coming from different sources (linked for example at territorial level)

Federated query on Istat and ISPRA, i.e. the query accesses Istat and ISPRA portals

Query on one Portal

Results dynamically retrieved from both portals

ISPRA - The Italian National Institute for Environmental Protection and Research

UN World Data Forum, Cape Town, 15 -18 January 2017
Example Query:
Municipality-level analysis of soil consumption and number of buildings by period of construction

ISTAT data: Census Buildings
ISPRA data: Data on soil consumption

Dynamically generated!
Use Case 3: Istat as Open Data Provider in SPOD - 1

SPOD: Social Platform for Open Data
RooteToPA H2020 Research and Innovation Project

Census LOD Rest

The webservice CensusLodREST allows access to triplestore Istat and get the dataset of the census variables aggregated by territorial sub municipal partitions.

The webservice responds at the endpoint http://datiopen.istat.it/CensusLodREST/webresources/service/getVariabiliCensuarie and takes the following parameters:

- **regione**: region
- **comune**: municipality
- **territorio**: territorial unit
- **dataset**: dataset
- **formato**: format
- **lang**: language

Parameter admitted values list:

<table>
<thead>
<tr>
<th>regione</th>
<th>comune</th>
<th>territorio</th>
<th>dataset</th>
<th>formato</th>
<th>lang</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lazio</td>
<td>Roma</td>
<td>areesubcomunali</td>
<td>popolazioneresidente</td>
<td>json</td>
<td></td>
</tr>
</tbody>
</table>


Dynamic Integration of a social platform with Istat’s data

UN World Data Forum, Cape Town, 15-18 January 2017
Use Case 3: Istat as Open Data Provider in SPOD - 2

Social discussion (on the left) about a graphic representation of Census Data (on the right)
Conclusions

- A dissemination strategy based on open data does put the Official Statistics users at the center:
  - Reaching them through different channels
    - e.g. apps and social media
  - Making easier for them to retrieve data
    - e.g. federated query that make transparent the distribution of data on different portals
  - Providing reacher services to them
    - e.g. spatial querying and dynamical visualizations

Open data and in particular Linked Open Data have a leading role in data innovation for Official Statistics