

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

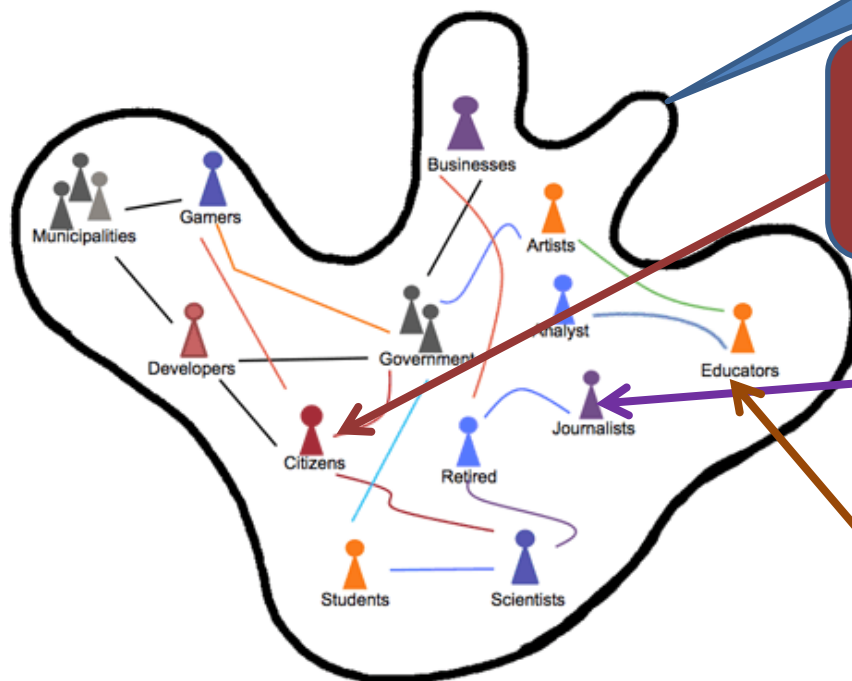
Giorgio Alleva
President

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.”*
(McKinsey Global Institute Report, Oct. 2013)

An Open Data Ecosystem



Wide Range of Open Data Consumers

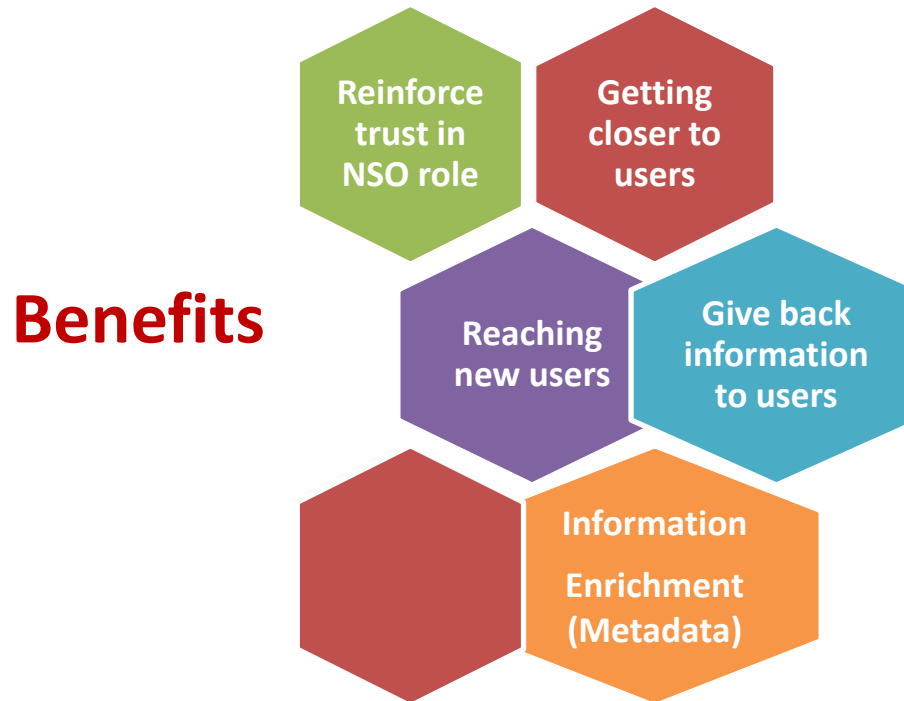
Citizens that would like to learn characteristics of the places they are, e.g. with mobile apps showing location-specific features

Journalists that need to access data for updated and aware information communication

Educators that are helped in their teaching task by access to data on different application domains

From Citizens IBM BLOG

- 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**



Semantic Web
Technological
Standards



Knowledge
Representation



Linked Open Data - LOD



Istat's Linked Open Data Portal - 1

The National Institute of Statistics | Tips | About | Census Variables Webservice ▾ | Browse data ▾ | ITA | ENG

linkedopendata | PIATTAFORMA SPERIMENTALE
PER I LOD DELL'ISTITUTO NAZIONALE DI STATISTICA

Vai alla pagina iniziale



Dati del Censimento della popolazione e delle abitazioni 2011

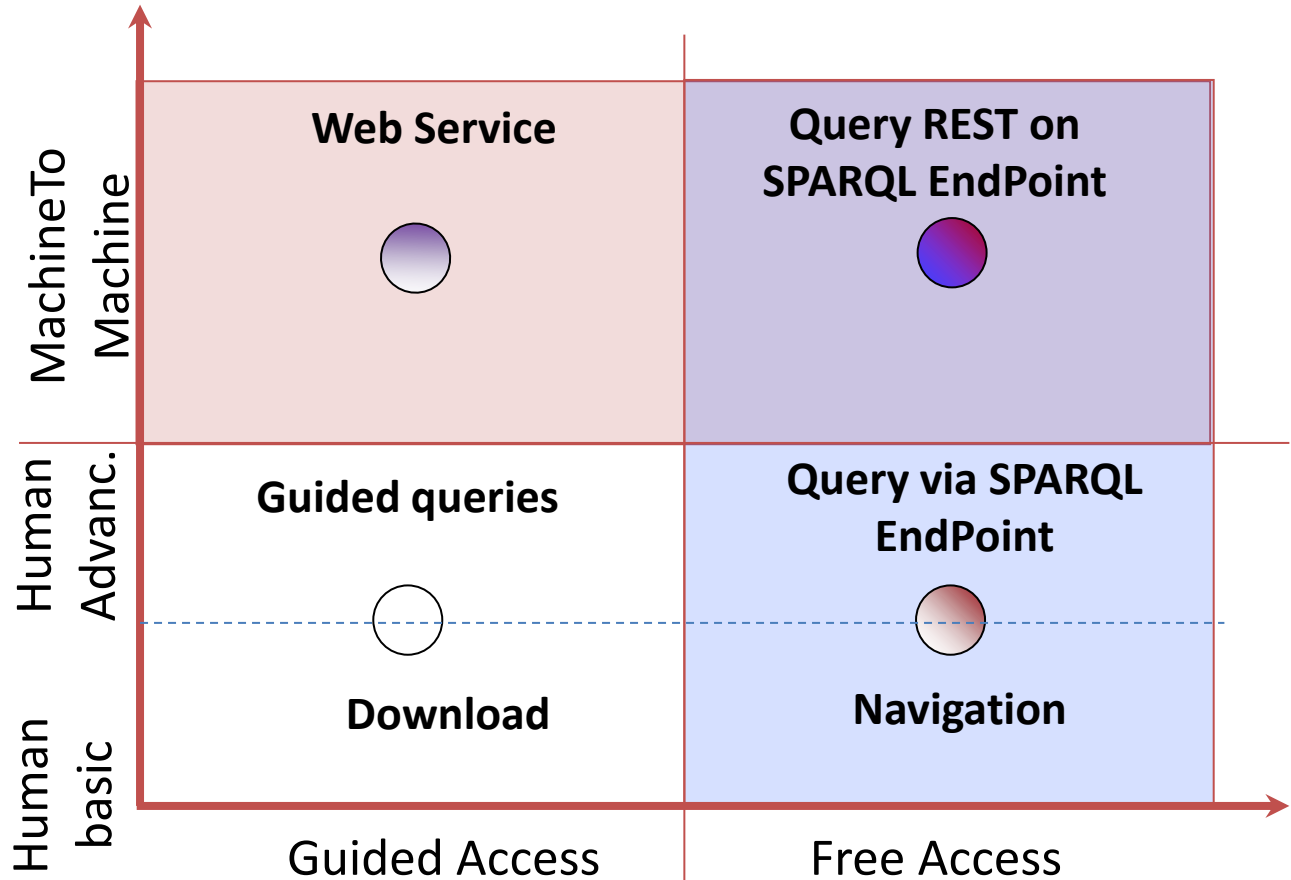
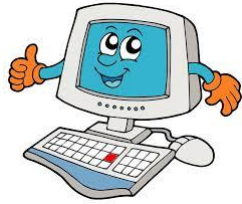
The Linked Open Data (LOD) portal allows to access and browse data in open format based on semantic web technologies and standards. The LOD portal can be queried directly from any application and responds to the need, expressed by the community of users, to have standardized and interoperable data.

TERRITORIAL LEVEL Wizard queries: territorial data year 2011	CENSUS VARIABLES Wizard queries: Aggregation by territorial sub municipal partitions Aggregation by Labour Market Areas	SPARQL ENDPOINT Custom queries: direct access to ENDPOINT SPARQL	DATASET/ONTOLOGIES Download census dataset Download Territorial and census data ontology
--	---	--	---

linkedopendata | Contact us | Privacy | Legal notice | WebInfo

Istat LOD Portal: <http://datiopen.istat.it>
English Version: <http://datiopen.istat.it/index.php?language=eng>

Interaction



Access freedom



Use Case 1: Spatial Querying

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

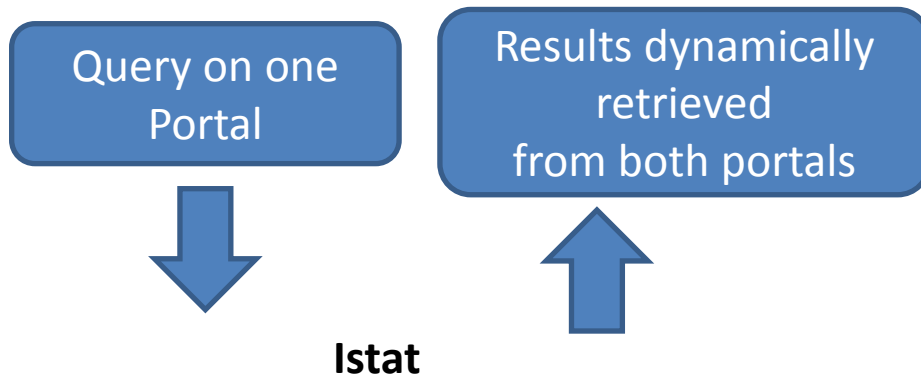
App that displays on a map some population indicators of the nearest census sections to specific GPS coordinates



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



ISPRA - The Italian National Institute for Environmental Protection and Research

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.

- TERRITORIAL LEVEL**
Wizard queries: territorial data year 2011
- CENSUS VARIABLES**
Wizard queries: Aggregation by territorial sub municipal partitions, Aggregation by Labour Market Areas
- SPARQL ENDPOINT**
Custom queries, direct access to ENDPOINT SPARQL
- DATASET/ONTOLOGIES**
Download census dataset, Download Territorial and census data ontology

LINKEDISPRA
I Linked Open Data dell'Istituto Superiore per la Protezione e la Ricerca Ambientale

417.845 entità / 3.728.447 triple

- ReNDiS: GLI INTERVENTI PER LA DIFESA DEL SUOLO
- CdS: IL CONSUMO DI SUOLO IN ITALIA
- RMN: LA RETE MAREOGRAFICA NAZIONALE
- RON: LA RETE ONDAMETRICA NAZIONALE

Use Case 2: Federated Querying - 2

Example Query:

Municipality-level analysis of soil consumption and number of buildings by period of construction

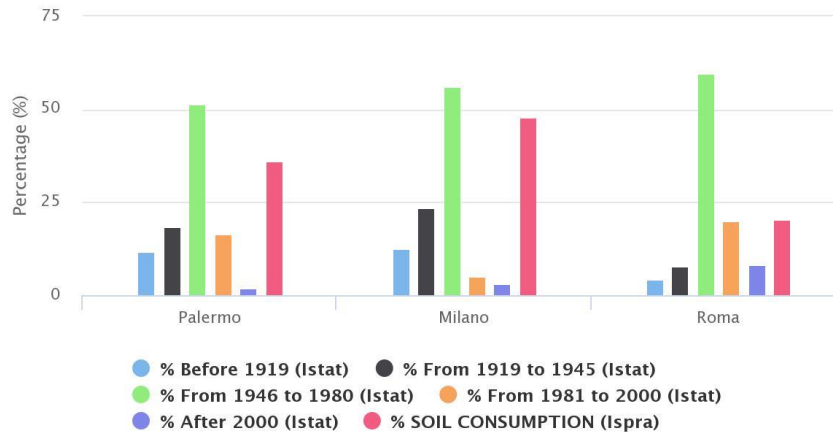
ISTAT data:
Census Buildings



ISPRA data:
Data on soil consumption

Residential buildings by period of construction and Soil consumption

Source: Istat - Ispra



Dynamically generated!

Highcharts.com

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



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

Dynamic Integration of a social platform with Istat's data

Census LOD Rest

linkedopendata

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 comune territorio dataset formato lang

Parameter admitted values list:

regione	comune	territorio	dataset	formato	lang
Link to query sparql	Link to query sparql	sezionidicensimento	popolazioneresidente	csv	it
		localita	popolazionestraniera	tsv	en
		areedicensimento	famiglie	text	
		areesubcomunali	gradodiistruzione	json	

<http://datiopen.istat.it/CensusLodREST/webresources/service/getVariabiliCensuarie?regione=Lazio&comune=Roma&territorio=areesubcomunali&dataset=popolazioneresidente&formato=json>

Use Case 3: Istat as Open Data Provider in SPOD - 2

The screenshot shows a social media feed titled "Popolazione 2011". The feed contains several posts from "Platform Administrator" and "UserC". The main post is a bar chart titled "Popolazione di Roma per fascia di età" (Population of Rome by age group). The chart shows population counts for various age groups, with the largest group being "Fascia Età Maggiore di 74 Anni" at 279,837. The chart is dynamically generated, as indicated by the "SUM(pop)" legend and the "Powered by ROUTE-TO-PA" logo. A source link is provided: <http://datiopen.istat.it>. The social discussion on the left includes comments like "Molto interessante" and "Sono d'accordo con il grafico".

Fascia Età	Popolazione
Fascia Età Minore di 5 Anni	120 046
Fascia Età Compresa Tra 5 e 9 Anni	118 017
Fascia Età Compresa Tra 10 e 14 Anni	113 639
Fascia Età Compresa Tra 15 e 19 Anni	115 105
Fascia Età Compresa Tra 20 e 24 Anni	119 648
Fascia Età Compresa Tra 25 e 29 Anni	131 968
Fascia Età Compresa Tra 30 e 34 Anni	159 297
Fascia Età Compresa Tra 35 e 39 Anni	206 628
Fascia Età Compresa Tra 40 e 44 Anni	222 579
Fascia Età Compresa Tra 45 e 49 Anni	223 870
Fascia Età Compresa Tra 50 e 54 Anni	188 038
Fascia Età Compresa Tra 55 e 59 Anni	164 906
Fascia Età Compresa Tra 60 e 64 Anni	161 461
Fascia Età Compresa Tra 65 e 69 Anni	144 532
Fascia Età Compresa Tra 70 e 74 Anni	147 604
Fascia Età Maggiore di 74 Anni	279.837

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

Dynamically generated!

- 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