

KDI

Semantic Heterogeneity Examples

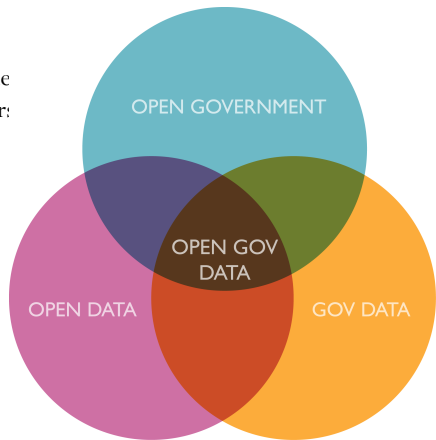
Fausto Giunchiglia and Mattia Fumagalli

University of Trento



Open Data Projects

Open data is data that anyone can access, use or share. Simple as that. When big companies or governments release non-personal data, it enable small businesses, citizens and medical researcher: to develop resources which make crucial improvements to their communities.



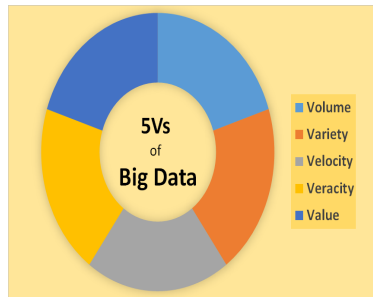
Open Government Data

- **Various governmental departments as part of their daily activities, produce, manage and store large volume of authentic and interesting data**
- **Why opening data?:**
 - great economic value
 - strong potential for supporting innovation
 - transparency and participation
 - improving organizational and communication efficiency
 - support data-centric applications
- **Not all of this data can be made publicly available because of the constraints such as:**
 - privacy issues
 - intellectual property rights
 - national security concerns



Open Data Vs. Big Data

Open in Open Data is about the modality by which data are made available. Big in big data is about the 5v which define the characteristics of data which are made available. The definition of Big Data as provided by Gartner is “*Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*”.



Open Research Data in HORIZON 2020



HORIZON 2020 GRANTEES ARE REQUIRED

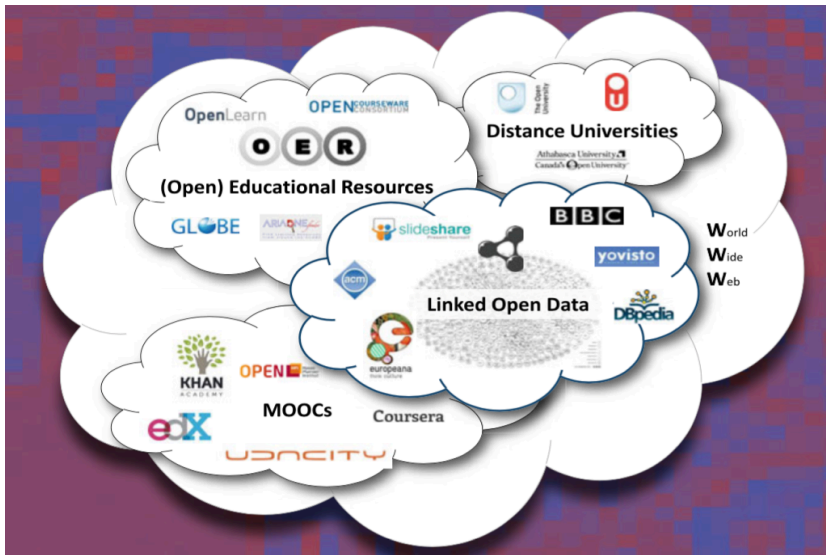
take measures to ensure open access to the **data** underlying their scientific publications

provide open access to **any other research data** of their choice

Horizon 2020 grantees are **encouraged to also share datasets** beyond publication



Open Data In Education



Famous Open Data Projects

GeoNames provide a rich source of information for geographical data. It contains millions of geographical names and unique features, such as mountain, city and airport. GeoNames provides spatial properties (e.g. longitude, latitude) as well as non-spatial properties (e.g. population) of spatial entities, but it lacks in support for accommodating diverse data sources.



OpenStreetMap (OSM)

OpenStreetMap (OSM) is a collaborative project to create a free editable map of the world. Two major driving forces behind the establishment and growth of OSM have been restrictions on use or availability of map information across much of the world and the advent of inexpensive portable satellite navigation devices. OSM is considered a prominent example of volunteered geographic information. --Wikipedia



LinkedGeoData (LGD) is the RDF (resource description framework) data dump of OpenStreetMaps (OSM) entities. Concept disambiguation (i.e., diversity in the nomenclature) and ontological correctness of the model have been overlooked in this geospatial resource.



Wiktionary (whose name is a blend of the words wiki and dictionary) is a multilingual, web-based project to create a free content dictionary of all words in all languages. It is available in 172 languages and in Simple English. It is written collaboratively by volunteers, dubbed "Wiktionarians". Its wiki software, MediaWiki, allows almost anyone with access to the website to create and edit entries



Free. The data in Wikidata is published under the Creative Commons Public Domain Dedication 1.0, allowing the reuse of the data in many different scenarios.

Collaborative. The data in Wikidata is entered and maintained by Wikidata editors, who decide on the rules of content creation and management in Wikidata.

Multilingual. Editing, consuming, browsing, and reusing the data is fully multilingual.



Wikipedia is a free encyclopedia, written collaboratively by the people who use it. It is a special type of website designed to make collaboration easy, called a wiki. Many people are constantly improving Wikipedia, making thousands of changes per hour.



DBpedia (from "DB" for "database") is a project aiming to extract structured content from the information created as part of the Wikipedia project. This structured information is then made available on the World Wide Web



Freebase was a large collaborative knowledge base consisting of data composed mainly by its community members. It contains data harvested from several sources such as Wikipedia, MusicBrainz, as well as individually contributed data from its users.

The screenshot shows the Freebase web interface for the entity 'Berlin'. At the top, there is a search bar with 'Berlin' entered and navigation links for 'Browse', 'Query', and 'Help'. A 'Sign In or Sign Up' link and the language 'English' are also visible. Below the search bar, the entity 'Berlin' is displayed with a thumbnail image of the Reichstag building at night. To the right of the image, there is a 'Cited by' section listing sources like 'en:imdb500', 'en:wikidata:Berlin', and 'en:the world:Berlin'. A large text block provides a detailed description of Berlin as the capital of Germany, mentioning its population of 3.5 million, its location on the River Spree, and its historical significance as the capital of both East and West Germany. Below the description, there are tabs for 'Properties', 'rdfs', 'Keys', and 'Links'. A filter bar allows users to view and edit specific domains, types, or properties. On the right side, a 'Types' sidebar lists various categories such as 'Common Type', 'Film', 'Filing location', 'Government', 'Location', 'City/Town/Village', 'Location', 'Administrative Division', 'German state', 'German city', and 'Statistical region'. The 'Description' tab is currently selected, showing a detailed paragraph about Berlin's status as the capital of Germany and its historical context.

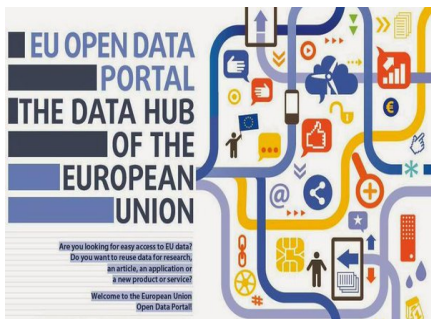
YAGO is a huge semantic knowledge base, derived from Wikipedia WordNet and GeoNames. Currently, YAGO has knowledge of more than 10 million entities (like persons, organizations, cities, etc.) and contains more than 120 million facts about these entities.



Open Data Sources

EU open Data Portal

The European Data Portal harvests the metadata of Public Sector Information available on public data portals across European countries. Information regarding the provision of data and the benefits of re-using data is also included.



Browse Categories in EU Open Data

Browse Categories



Agriculture,
Fisheries,
Forestry &
Foods



Energy



Regions &
Cities



Transport



Economy &
Finance



International
Issues



Government
& Public
Sector



Justice, Legal
System &
Public Safety



Environment



Education,
Culture &



Health



Population &
Society



Science &
Technology

Data domains with the most datasets

| Copenhagen | London | Paris |
|------------------------|---------------|-------------------|
| Geospatial | Demographics | Transportation |
| Transportation | Employment | Administration |
| Children & adolescents | Health | Culture |
| Statistics | Transparency | Urban development |
| | Housing | |

Transportation is among the most interesting data categories for all the cities included.

Open data trentino portal harvests the metadata of Public Sector Information available on public data portals in Trentino region.



CKAN is a powerful data management system that makes data **accessible** – by providing tools to streamline **publishing**, sharing, **finding** and **using** data. CKAN is aimed at data publishers (national and regional governments, companies and organizations) wanting to make their data open and available.



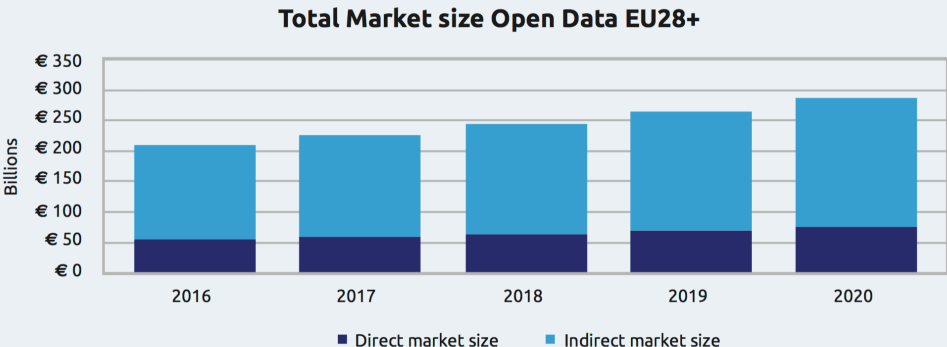


Figure – Total market size (high bound), split in direct and indirect size for EU28+ in billions, 2016-2020

Direct market size per sector

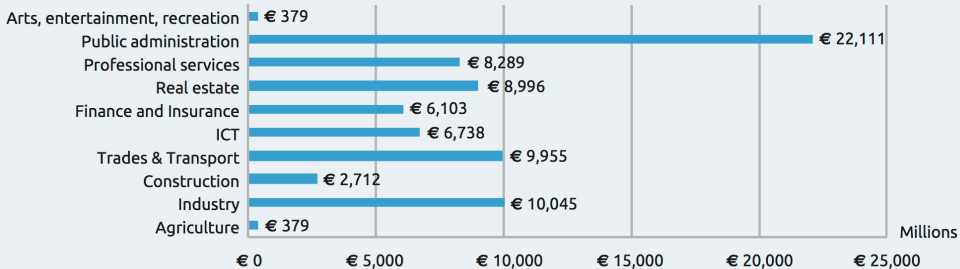


Figure – Direct market size of Open Data per market sector for EU28+ in millions, 2020

Benefit of Open Data

more Open Data can help make

better decisions



2,549 hours
wasted
finding parking



7,000 lives

saved due to
quicker response



629 million

hours saved is
equivalent to

€ 27.9 bn

5.5% less
road fatalities



Congestion

costs are
1% of GDP



16% less
less energy used

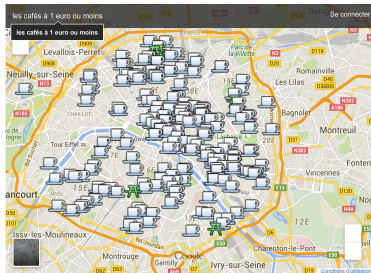


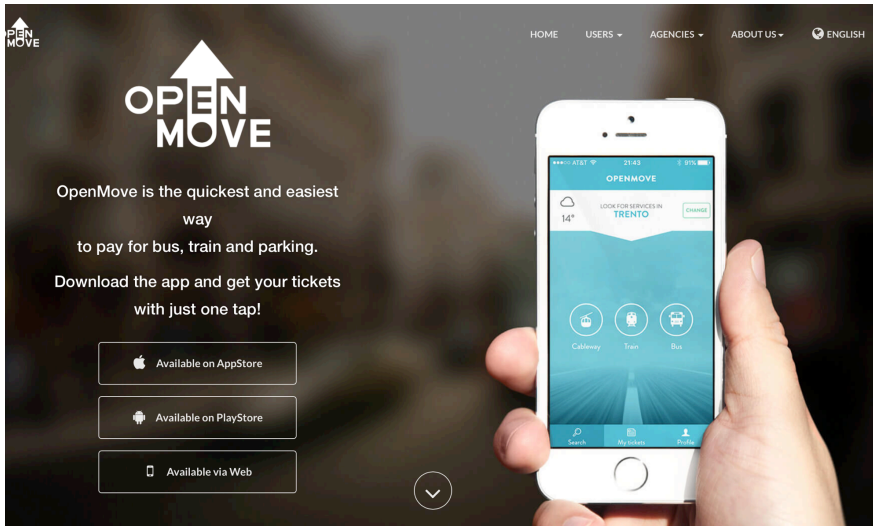
$$\text{Benefit} = \frac{\text{Cost savings} + \text{Increased Returns}}{\text{Costs}}$$

Innovation and growth in businesses

Open data is supporting innovation and growth by revealing opportunities for businesses large and small to build new services, identify savings and improve operations.

In Paris, café owners are using open data to attract tourists to the city's most affordable coffees, while, in Iceland, farmers are using open data about the quality of their lambs to attract new customers





OPEN MOVE

HOME USERS AGENCIES ABOUT US ENGLISH

OPEN MOVE

OpenMove is the quickest and easiest way to pay for bus, train and parking.

Download the app and get your tickets with just one tap!

Available on AppStore

Available on PlayStore

Available via Web

OPENMOVE
14°
LOOK FOR SERVICES IN TRENTO
CHANGE
Cableway Train Bus
Search My tickets Profile

↓

Benefiting culture and the environment

Open data helps farmers to improve yields and support a growing population without the need to destroy valuable habitats.

Plantwise are collecting open data to produce valuable information packs for farmers about plant health and threats from diseases.



