

Linked Open Data for Smart Regions and Citizens

Lorenzo Sommaruga¹, Nadia Catenazzi¹, Riccardo Mazza¹ and Davide Bertacco¹

¹ *University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Polo universitario Lugano - Campus Est, Via la Santa 1, CH-6962 Lugano – Viganello, Switzerland*

Abstract

A lot of data useful for citizens are stored in isolated silos and proprietary formats. Their publication in a homogeneous open standard format, as Linked Open Data (LOD), can provide a number of potential benefits, accelerating the creation of smart cities and regions, with added value services. In order to facilitate the LOD production, the GIOCOOnDa LOD platform is introduced, mainly focusing on the data publishing process. Then, the paper focuses on data consuming, presenting two examples of applications developed using the GIOCOOnDa LOD datasets. Finally, a vision of smart cities and regions derived from LOD and smart data is introduced.

Keywords

Open Data, LOD, publication, interlinking, smart citizens, smart regions, smart data

1. Introduction

A huge amount of data, potentially useful for everyday citizens' life, is being generated by both public and private organizations on an ongoing basis. These data are often stored in government or proprietary databases and maintained isolated in data silos according to the organization needs; they are not based on a common data model and no connections are explicitly declared to other related datasets. This leads to a number of issues, including inconsistency, duplication, waste of resources, and inability to get a comprehensive view of data towards a vision of common good.

Opening data solves part of these issues, providing a number of potential benefits, such as improved efficiency of public administrations, innovation and economic growth, public service improvements [1, 2].

The proliferation of open data, i.e., “data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike” [3], will certainly provide further opportunities for ICT to make inroads into many application domains.

Smart environments, in general, and smart cities, in particular, are certainly promising fields also from the business point of view as introduced in [4]: “Smart cities need open data. Smart cities are being designed to improve the quality of life of individuals and communities – from transport and education to water, waste and energy. These smart infrastructures demand an open culture that will benefit everyone”.

Open data could help to accelerate the development of smart cities by acting as enablers. At this purpose, open data need to be reliable, accurate and accessible. In this way they can provide key data sources to people with technical knowledge that can develop citizen-centric solutions: “In smart cities, open data represents a powerful tool to facilitate open innovation and to foster urban innovation” [5].

Proceedings of FTAL 2021, October 28–29, 2021, Lugano, Switzerland

EMAIL: lorenzo.sommaruga@supsi.ch (A. 1); nadia.catenazzi@supsi.ch (A. 2); riccardo.mazza@supsi.ch (A. 3); davide.bertacco@supsi.ch (A. 4)

ORCID: 0000-0003-0254-4952 (A. 1); 0000-0001-6029-2138 (A. 2); 0000-0002-9074-6927 (A. 3)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Furthermore, in order for data to be truly open, it is useful to follow the Findable, Accessible, Interoperable and Reusable (aka FAIR) data principles, already well promoted by national programmes and adopted by the research community [6]. The FAIR approach may contribute to accelerate innovation also in the public administration and business communities, for making data more findable, accessible, interoperable, and reusable. This will certainly contribute, in the next decades, to pose the basis for providing an effective tool to facilitate urban innovation.

An example of application of open data for smart cities is shown by the city of Brussels, where openness of data will make life easier for residents, providing access to town services and contributing to the government transparency. Through their *Opendatastore* and Open Data Portal platform, using the CKAN standard solution, public administrations and other stakeholders can publish open data in common open data formats, such as CSV, GML, JSON or ZIP. These data can be easily accessed and reused by citizens and developers for their own purposes, for instance, providing tourists with real-time information on ATMs, parcours, and others. In addition, anonymized data collected and analyzed from a public mobile network, allow administrators to promptly make better decisions on allocating resources [4].

A step further can be accomplished with Linked Open Data (LOD), the fifth star level in the Berners' Lee model [7]; LOD technologies provide an infrastructure for data integration and merging, where data can be accessible via standard web technologies. The potential impact derived from the application of linked data principles is enormous: data are interoperable, reusable, new knowledge can be automatically derived from data merging. In addition, the use of linked data reduces managing costs (once the basic infrastructure has been set up) and adds a new value derived from data integration. Thanks to these features, LOD is the basis for smart data, i.e., data that makes itself understandable, extracting relevant insights from large data sources and presenting the conclusions as human-friendly visualizations [8].

An interesting Swiss linked open data initiative is represented by the Linking Geodata Service [9], which provides access to a number of datasets of the Federal Spatial Data Infrastructure (such as public transport stops, road traffic accidents and Swiss boundaries 3D), presents some SPARQL query examples and makes reference to linked open data portals, such as LINDAS [10], a new platform available since 2020 to make data available as linked data.

In general, setting up an infrastructure to publish linked open data is more complex than producing 2* or 3* open data. There is the need to model data according to standard ontologies and to create interconnections to other related datasets.

The GIOCOOnDa LOD platform provides an answer to this challenge, allowing the conversion from Open Data (3*) to Linked Open Data (5*). This platform has been developed in the context of the two-year EU Interreg GIOCOOnDa project (“Integrated and holistic management of the open data life cycle” [11]), dedicated to governance, whose general objective is to strengthen the coordination and collaboration capacities of Italian and Swiss public administrations to create responses to the lack and under-use of common information by of the territories. The GIOCOOnDa project has been partially funded by Interreg V A European Regional Development Fund, the Tessin Canton and the Swiss Confederation.

The project publishes linked open data about the Insubric region, a cross-border territory and community across Italy and Switzerland. These include data on museums, accommodation facilities and environment, taken from various data sources: Regione Lombardia open data portal [12] and ARPA (Regional Agency for the Protection of the Environment) [13] for Italian data; Wikidata, Ticino Turismo [14] and OASI [15] for Swiss data.

This paper is organized as follows: in Section II the GIOCOOnDa LOD platform is introduced, mainly focusing on the data publishing process; section III is focused on data consuming, presenting some examples of applications developed using the GIOCOOnDa LOD datasets; finally, section IV is dedicated to highlight the advantages of LOD data for smart citizens and regions.

2. Data Publishing: the GIOCOOnDa LOD Platform

A number of guidelines, best practices and recommendations have been produced to guide the process of Linked Open Data publishing [16, 17, 18]. The approach adopted in GIOCOOnDa follows the AGID guidelines [18] and consists of the following steps:

- *selection of dataset*: on the basis of the results of a previous need analysis, data about museums, accommodation facilities and environmental data are selected;
- *data cleaning*: these datasets are already published as “clean” open data, and a quality check is already accomplished;
- *analysis and RDF modelling*: datasets are analysed and appropriate ontologies are identified to model them. In particular, ontologies of the OntoPia network [19] are chosen. In the GIOCOOnDa LOD platform, data are imported from different sources and converted into RDF, according to these standard ontologies;
- *enrichment*: datasets are enriched with metadata, following the DCAT-AP standard;
- *interlinking*: data are interlinked to other datasets such as Wikidata;
- *publication*: datasets are published using Openlink Virtuoso Universal Server and can be queried through a SPARQL endpoint.

The conversion of heterogenous data into linked open data guarantees standardization and semantic interoperability. It is a complex process that is usually implemented by developing ad-hoc middleware and requires programming skills. In the GIOCOOnDa LOD platform, this process can be configured by an expert who is not a programmer, through a graphical user interface. The platform, implemented as a Java based web application, provides different functionalities that enable the publication of LOD datasets starting from open datasets, and their visualization in a catalogue or in a map. In addition, the platform allows a data provider to filter and select which data to make public in their dataset, hiding some sensitive data and exposing only the most relevant and legitimate ones, thus taking care for instance of privacy issues.

Further details about the GIOCOOnDa LOD platform are specified in [20], where the conversion process from open data to LOD is described step by step, by means of simple real examples, and the different functionalities implemented in the platform are presented. The resulted LOD datasets can be used to create *consuming* applications; two examples are here presented.

3. Data Consuming: examples of APPs

One of the main advantages of LOD is the possibility of having heterogenous data from more sources available in a uniform and accessible format via standard web technologies. This permits the implementation of applications that can exploit the enrichment and interlinking of LOD to provide new views of the data and derive new information.

We will illustrate this opportunity with two case studies that are implemented in the context of the GIOCOOnDa project: a unified view of Tessin and Lombard museums, whose data are already described in the previous section, and a list of Tessin and Lombard holiday accommodations.

3.1. Museums in Tessin and Lombardy

The GIOCOOnDa platform provides a SPARQL endpoint that can be queried to extract LOD data from the RDF data sources stored in the Virtuoso server. Thanks to the standard syntax of SPARQL, we are able to fetch homogeneous data to implement new applications, or to connect with third-part applications, without the need of knowing the complex notions of ontologies and RDF.

In the first case study, the idea is to implement an application that provides a list of museums in Tessin and Lombardy, along with their main information: name, type of museum, location, address, and a picture. The target users are people interested in knowing the various type of museums that exist in

these regions, how many museums are available for each typology and their distribution across the regions. The purpose of the application is hence to explore and navigate these data, that can be useful to take data-driven decisions, such as opening a new museum or establish a new partnership among museums covering the same subjects but located in different regions.

This application is based on linked open data about museums, previously converted and interlinked in the GIOCOOnDa platform. The original data sources are the Regione Lombardia portal and Wikidata.

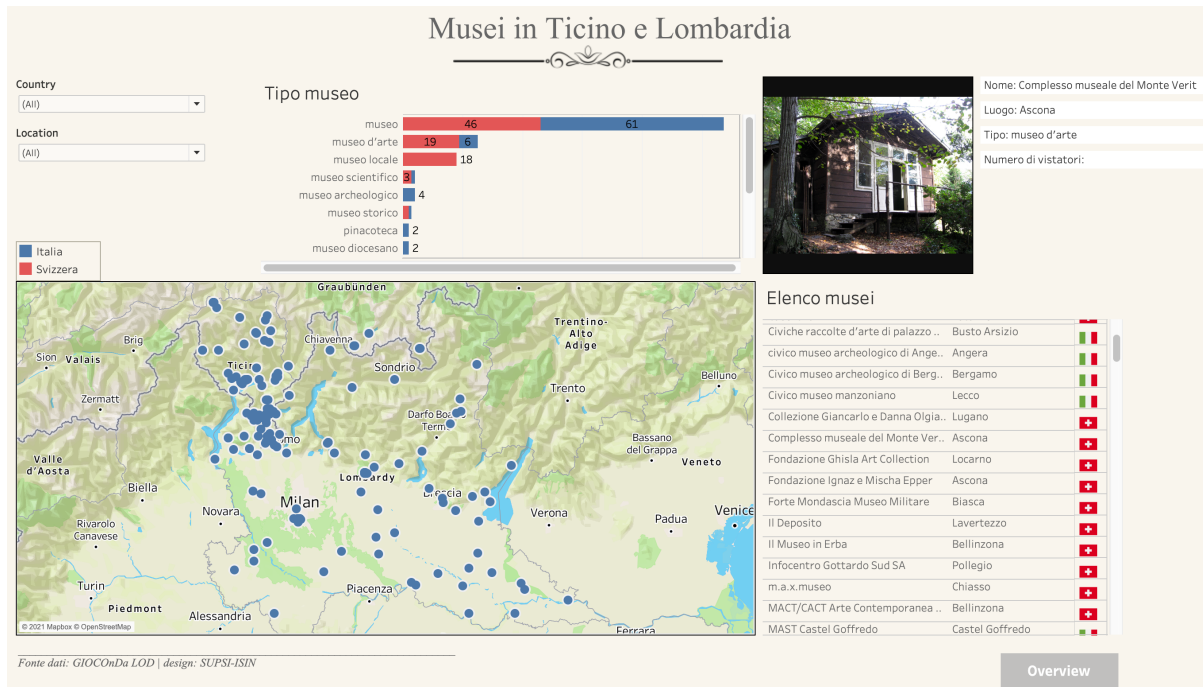


Figure 1: Application to explore museums

These data are extracted from the data.world platform [21] through the SPARQL interface. Data.world is a cloud-based service that allows anyone to create a data catalog that can be integrated with dozens of applications for analytics and visualization. We exploited the integration of Data.world with the visual analytics platform Tableau [22], and implemented the dashboard illustrated in Figure 1. Using this dashboard, the viewer is able to graphically represent LOD data extracted from GIOCOOnDa datasets, to make sense of the various types of museums and their spatial distribution in the two regions.

3.2. Holiday accommodations in Tessin and Lombardy

In another case study, we switched our focus to holiday accommodations. Again, the purpose of this application is to explore the different types of accommodation across the two regions in a uniform way. In that case, the source data come from Wikidata for Italian data and Ticino Turismo for the Swiss data. We used the same infrastructure described in the previous case study, but we obviously changed the SPARQL and visual interface to extract and visualize data that are relevant to our goal.

Figure 2 shows the dashboard implemented for this purpose. Different types of accommodation are encoded in the map with circles of different colours, while the bar chart on the top shows the types, partitioned (in quantity) between the two regions. A simple example of data-informed decisions that can be drawn from this visualization is the following: an entrepreneur could decide to open a farm house (“Agriturismo”) or a Bed & Breakfast in Tessin, which are highly underrepresented with respect to Lombardy.

Accommodations in Ticino and Lombardia

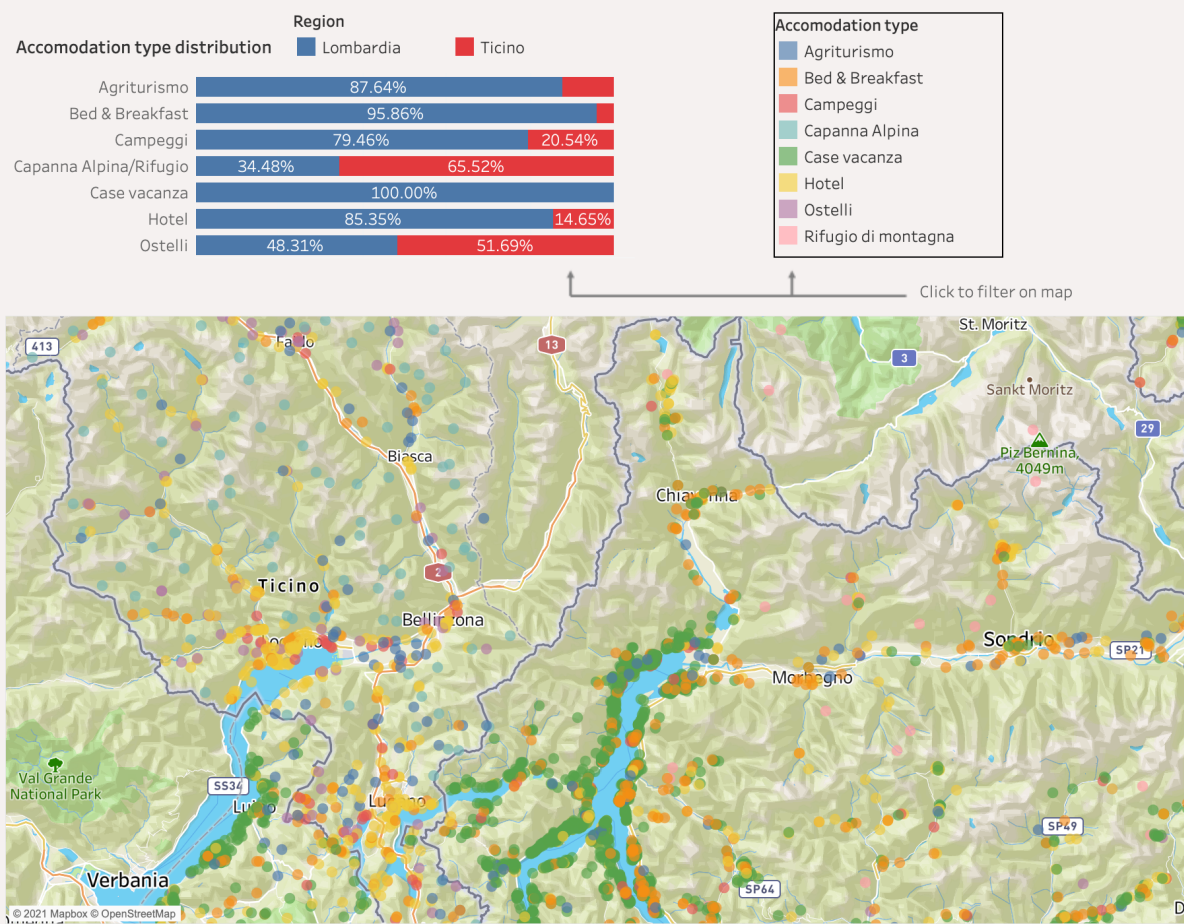


Figure 2: Application to explore holiday accommodations

Although the potential of linked open data and smart data for smart cities is widely recognized and the supporting technologies are available, there are still some drawbacks for their adoption. The main obstacle is the difficulty to generate LOD and their poor availability, as also confirmed by the analysis carried out in the GIOCONDa project: the publication of LOD by organizations is still not yet a common practice of the data management processes, and developments are relatively slow.

4. Conclusions

This paper has envisioned the potential of Linked Open Data for smart regions and presented the GIOCONDa LOD platform to publish LOD, as well as some applications that demonstrate their benefits for citizens. The GIOCONDa LOD platform facilitates the conversion of heterogeneous data to homogeneous filtered linked open data, which enable an easy generation of added value services.

The GIOCONDa project achieves a two-fold result: on one hand, it represents a small step towards the creation of a platform for LOD data, where datasets with potential impact on the territories can easily be made available to everyone, according to vocabularies and standard formats; on the other hand, it establishes a big step for the Insubric interregional area towards raising awareness about the advantages of opening public administration data. On this basis, it will now be possible to sow other data and reap the benefits of their use, especially with the development of apps that give vision and support to decisions, towards smarter regions and happier citizens [23].

5. Acknowledgements

We acknowledge all the partners involved in the GIOCOnda project for their contribution to the different project activities during 2019-2021.

6. References

- [1] data.europa.eu, What is open data. URL: <https://data.europa.eu/en/trening/what-open-data>, retrieved November 2021.
- [2] The World Bank, Starting an Open Data Initiative. URL: <http://opendat toolkit.worldbank.org/en/starting.html>, retrieved November 2021.
- [3] Open Knowledge Foundation, Open Data Handbook – What is Open Data”. URL: <https://opendatahandbook.org/guide/en/what-is-open-data/>, retrieved November 2021.
- [4] J. Howells, Smart cities need open data. URL: <https://www.orange-business.com/en/blogs/smart-cities-need-open-data>, March 12, 2018, retrieved November 2021.
- [5] L. Smith, Benefits of Open Data for Smart Cities, Bee Smart City. URL: <https://hub.beesmart.city/en/solutions/benefits-of-open-data-for-smart-cities>, Dec. 2017, retrieved November 2021.
- [6] FAIR, FAIR Principles. URL: <https://www.go-fair.org/fair-principles/>, retrieved November 2021.
- [7] T. Berners-Lee, Linked Data. URL: <https://www.w3.org/DesignIssues/LinkedData.html>, 2009, retrieved November 2021.
- [8] M. Emaldi, O. Peña, J. Lázaro, D. López-de-Ipiña, “Linked Open Data as the Fuel for Smarter Cities”, in: F. Xhafa, L. Barolli, A. Barolli, P. Papajorgji (Eds.), Modeling and Processing for Next-Generation Big-Data Technologies. Modeling and Optimization in Science and Technologies, Volume 4, pp. 443-472, Springer, Cham, 2015. URL: https://doi.org/10.1007/978-3-319-09177-8_18.
- [9] Geo.admin.ch – the federal geoportal, Linked Data Service: linking geodata. URL: <https://www.geo.admin.ch/en/geo-services/geo-services/linkedata.html>, retrieved November 2021.
- [10] LINDAS – Linked Data Service, LINDAS. URL: <https://lindas.admin.ch>, retrieved November 2021.
- [11] Interreg PROGETTI , GIOCOnda, Integrated and holistic management of the Open Data life cycle (Gestione integrata e olistica del ciclo di vita degli Open Data). URL: <https://progetti.interreg-italiasvizzera.eu/it/b/78/gestioneintegrataeolisticadelciclodivitadegliopendata>, retrieved November 2021.
- [12] Regione Lombardia, Open Data. URL: <https://dati.lombardia.it>, retrieved November 2021.
- [13] Arpa Lombardia. URL: <https://www.arpalombardia.it/>, retrieved November 2021.
- [14] Ticino Turismo. URL: <https://www.ticino.ch>, retrieved November 2021.
- [15] OASI, Osservatorio Ambientale della Svizzera italiana, Esplora dati. URL: <https://www.oasi.ti.ch>, retrieved November 2021.
- [16] F. Bauer, M. Kaltenböck, Linked Open Data: The Essentials - A Quick Start Guide for Decision Makers. Edition mono/monochrom, Vienna, Austria, 2012, ISBN: 978-3-902796-05-9. URL: <https://www.recep.org/LOD-the-Essentials.pdf>, retrieved November 2021.
- [17] W3C, Best Practices for Publishing Linked Data, W3C Working Group Note 09 January 2014. URL: <https://www.w3.org/TR/ld-bp/>, retrieved November 2021.
- [18] Agenzia per Italia Digitale, Guidelines for semantic interoperability through Linked Open Data (Linee Guida per Interoperabilità semantica attraverso i Linked Open Data). URL: http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf, 2012, retrieved November 2021.
- [19] Ontologie e Vocabolari Controllati. URL: <https://github.com/italia/daf-ontologie-vocabolari-controllati>, retrieved November 2021.
- [20] L. Sommaruga, N. Catenazzi, D. Bertacco, R. Mazza, From Open Data to Linked Open Data - The GIOCOnda LOD platform, in: Proceedings of ALLDATA2021, The Seventh International

Conference on Big Data, Small Data, Linked Data and Open Data, April 18-22, 2021 – IARIA, Porto, Portugal, pp. 10-15, ISBN: 978-1-61208-842-6.

[21] data.world, The Cloud-Native Data Catalog, <https://data.world>, retrieved: May, 2021.

[22] Tableau: Business Intelligence and Analytics Software. URL: <https://www.tableau.com/>, retrieved November 2021.

[23] La pandemia non ha fermato il progetto GIOCOnda, VareseNews. URL: <https://www.varesenews.it/2021/04/la-pandemia-non-fermato-progetto-gioconda/1325687/>, 2021, retrieved November 2021.