

Linked Open Data in the GIOCOOnDa LOD Platform

Lorenzo Sommaruga, Nadia Catenazzi, Davide Bertacco, Riccardo Mazza

Department of Innovative Technologies

University of Applied Sciences and Arts of Southern Switzerland (SUPSI)

CH-6962 Lugano, Switzerland

e-mail: lorenzo.sommaruga@supsi.ch

nadia.catenazzi@supsi.ch

davide.bertacco@supsi.ch

riccardo.mazza@supsi.ch

Abstract – The GIOCOOnDa LOD platform, developed in the context of a project funded by the EU programme Interreg, aims to make Linked Open Data (LOD) available in an easy and convenient way, without the need of any software programming. This paper illustrates the implemented platform that, starting from a common Open Data repository, can automate the production and publishing of Linked Open Data. The platform is configurable and extensible as it enables to define mapping configurations for new datasets hiding the complexity of the ontology in the mapping process. The paper also presents some examples of data consuming applications, using some GIOCOOnDa LOD datasets.

Keywords - *Linked Open Data (LOD); GIOCOOnDa; LOD publishing; LOD consuming; OntoPia.*

I. INTRODUCTION

The work described in this paper, firstly introduced in [1], was developed in the context of the EU Interreg GIOCOOnDa project (“Integrated and holistic management of the open data life cycle”, March 2019 – April 2021), funded by the Interreg V-A Italy-Switzerland Programme, which aims to create value by developing information products based on the re-use of public Open Data [2]. The project focuses on data that are relevant for the touristic sector, coming from the Insubric area, a cross-border territory across Italy and Switzerland. They include data about museums, accommodation facilities, and environment. The main data sources currently used are: Regione Lombardia open data portal [3], and ARPA (Regional Agency for the Protection of the Environment) [4] for Italian data; Wikidata, Ticino Turismo [5], and OASI [6] for Swiss data.

One of the main results of the project is the creation of a platform for the publication of LOD (Linked Open Data) by public administrations. In the GIOCOOnDa LOD platform, open data, coming from various sources and in different formats, are converted into homogeneous Linked Open Data, based on standard ontologies, and published together with their metadata. The platform enables conversion of existing 3* Open Data to 5* Open Data, according to the well-known 5-star deployment scheme [7]: data are formalized in RDF, identified by URI and linked to other datasets. For this purpose, a specific interlinking module was developed and integrated into the platform.

This work answers an emerging need for generic and usable tools to produce linked open data. While the value of linked open data is widely recognized in the literature [8], their publication is still challenging.

Different works and platforms have been developed to support this process. One of the first significant projects is Lucero and the resulting Tabloid toolkit, which aims to help institutions and developers to publish and consume linked data [9]. Another interesting work, which supports US open government data production and consumption, is the TWC LOGD portal [10]. Here a workflow for linked open data deployment is defined, consisting of different stages, where the conversion process is automated by using the `csv2rdf4lod` tool. A more recent initiative is represented by the Italian cultural heritage platform “`dati.beniculturali.it`”, promoted by the Italian Ministry of Culture, which collects and publishes standardized and interoperable LOD heterogeneous datasets [11].

An extensive survey of methods, tools, and techniques for generating and publishing linked open data reveals that the proposed approaches to produce linked data are often specific to a use case and usually concern a specific domain, such as, media, library, finance, education, and healthcare. These approaches can hardly be adapted to other use-cases and domains [12].

The main innovative aspect of the approach adopted in the GIOCOOnDa platform is that it is generic and supports different source types and formats. In addition, it does not require programming skills or a deep knowledge of the RDF and OWL formalisms.

This paper describes the publishing process from Open Data to Linked Open Data in the GIOCOOnDa platform and presents some examples of applications that consume LOD produced using the platform. It is structured as follows: Section II presents the methodology adopted to publish LOD data; Section III focuses on the process of conversion of Open Data to LOD, one of the main steps of this methodology; Section IV presents the main functionalities of the GIOCOOnDa LOD platform, in particular the LOD catalogue, and the input and output mapping, two processes that enable to convert data from different input data sources to the final LOD format; Section V explains how the interlinking module works and is integrated into the platform; Section VI presents possible use cases of

publishing a new dataset in the GIOCONDa LOD platform; finally, Section VII proposes some examples of applications that exploit Linked Open Data to support users in their needs.

II. METHODOLOGY TO PUBLISH LINKED OPEN DATA

From a methodological point of view, a number of best practices, recommendations and guidelines have been defined. For example, Bauer and Kaltenböck [13] provide a step-by-step model, highlighting the most important issues that need to be considered in LOD publishing; W3C [14] presents best practices designed to facilitate LOD development and delivery; the “Agenzia per l’Italia Digitale” (AGID) [15] proposes a general methodological approach for the interoperable opening of public data through the LODs. This methodology basically consists of the following steps: selection of dataset, data cleaning, analysis and RDF modelling, enrichment, interlinking, validation, and publication.

The approach adopted in GIOCONDa is in line with the above best practices and guidelines and, in particular, with those proposed by AGID. The selection of datasets was made on the basis of the results of a previous need analysis phase carried out with a number of stakeholders during the project. As a starting point, data about museums, accommodation facilities, and environment of the Insubric region are selected.

Concerning data cleaning, it is assumed that the selected datasets are already published as *clean* and *accurate* open data, where a quality check is already accomplished.

Once selected, datasets are deeply analysed to understand their structure; then appropriate ontologies and vocabularies are identified to model them.

In particular, the adopted ontologies are taken from the OntoPia network [16], also presented in [17]. They include, for instance, the Cultural-ON ontology for museums and the ACCO ontology for accommodations. In the GIOCONDa LOD platform, data are imported from different sources and converted into the RDF format, according to these standard ontologies. The conversion process is detailed in the next section.

As additional steps, datasets are enriched with metadata and interlinked to other datasets. Metadata are added to the single datasets following the DCAT-AP standard. Interlinks to other datasets are created by identifying alignments and similarities between different datasets. For instance, a museum from the “Regione Lombardia” dataset can be declared “the same as” a museum described in Wikidata. The identification of interlinks is mainly carried out using the Silk software libraries [18], as explained in Section 5.

Finally, datasets are published using Openlink Virtuoso Universal Server [19], where they can be queried through a SPARQL endpoint.

III. THE CONVERSION PROCESS FROM OPEN DATA TO LOD IN THE GIOCONDA PLATFORM

The core of the system lies in the mapping functionality of heterogeneous data into Linked Open Data, according to standard ontologies.

This conversion is a complex process that depends on the initial format and on the final standard RDF format. From a

literature study it emerges that the most frequently adopted approach is the implementation of ad-hoc middleware. For example, to convert a relational database to LOD, a typical solution is to use declarative languages, such as D2R [20] or R2RML [21] that require ontological and programming skills.

In the GIOCONDa LOD platform, the complexity of the conversion process is simplified by defining a converter, facilitated by a graphical user interface that an expert can use to configure the conversion. This process can be explained through a simple example: we would like to convert two different datasets about museums into a common interoperable format. The first dataset concerns *Lombard museums* retrieved from the Regione Lombardia portal in JSON format by means of REST APIs [21]. The second is represented by *Tessin Canton museums* retrieved from Wikidata through SPARQL queries.

Figure 1 shows an excerpt from the Lombard museums visualized on the Regione Lombardia portal [22], while Figure 2 shows an example of a Swiss museum in Wikidata [23].

CODI...	CODI...	PROV...	COMUN...	DENOMINAZIONE_MUSEO	DENOMINAZIONE_SEDE
25	121	BG	BERGAMO	ORTO BOTANICO DI BERGAMO LORENZO ROTA	Orto Botanico di Bergamo 'Lor
2590	2770	MI	MILANO	POLO DEI MUSEI SCIENTIFICI	Acquario e civica Stazione Idro
421	2405	SO	CHIAVENNA	MUSEO DEL TESORO	Battistero
1888	1950	BG	LOVERE	MUSEO CIVICO DI SCIENZE NATURALI "ALESSIO ...	Sede espositiva
2461	2497	MI	CINISELLO B.L.	MUSEO DI FOTOGRAFIA CONTEMPORANEA	Museo di Fotografia Contemp
1982	2062	BG	TREVIGLIO	MUSEO CIVICO ERNESTO E TERESA DELLA TORRE	Museo Civico Ernesto e Teresa
2175	2217	MI	MILANO	Museo Nazionale della Scienza e della Tecnolog...	Museo Nazionale della Scienza
2204	2350	PV	PAVIA	SISTEMA Museo Nazionale della Scienza e della Tecnologia Leonardo da Vinci	Museo Nazionale della Scienza e della Tecnologia Leonardo da Vinci

Figure 1. Lombard museums from the Regione Lombardia Open Data portal.

Language	Label	Description	Also known as
English	Museo Vincenzo Vela	museum in Mendrisio (Switzerland)	
Italian	Museo Vincenzo Vela	museo a Mendrisio, Svizzera	Museo Vela
French	No label defined	musée en Suisse	
Sardinian	No label defined	No description defined	

Figure 2. The Swiss *Vela* Museum in Wikidata.

To be able to configure the mapping from the original to LOD format, the structure of the two museum data sources has to be analysed by an expert and an appropriate ontology selected. In this phase it is important to find the most appropriate ontology to model the domain. Cultural-ON [24]

and its connected ontologies were chosen because they are representative of the museum domain and can be exploited to support transnational interoperability.

The next step consists of analysing the different descriptive fields of the museum datasets: for instance, each Lombard museum is described in terms of 79 fields, such as *denominazione museo* (name), *telefono* (telephone), *codice sede* (site code) as shown in Figure 1.

For each field, the objective is to find a match with the ontology classes and properties. For example, a museum could be represented as an instance of the *cis:Museum* class of the Cultural-ON ontology, where *cis* is the prefix of the ontology namespace; the *telefono* field can be mapped into a property of a *smapit:OnlineContactPoint* instance of the Social Media / Contact and Internet ontology [25].

Figures 3 and 4 present some details of the conversion result of a Lombard and a Tessin canton museum, respectively, into RDF Turtle, according to Cultural-ON and its connected ontologies. In particular, in the excerpts, the light blue border highlights the *hasSite* relation to the *Site* instance, and the orange one highlights the *hasOnlineContactPoint* relation to the *ContactPoint* instance with their respective properties.

```

museum:Museo_2175_sede_2217_Museo_..._Leonardo_da_Vinci
a cis:CulturalInstituteOrSite, cis:Museum ;
rdfs:label "Museo Nazionale ... Leonardo da Vinci" ;
cis:institutionalName "Museo ... Leonardo da Vinci" ;
cis:hasSite site:Sede_2217 ;
smapit:hasOnlineContactPoint
contactPoint:Contatti_Museo_Leonardo_da_Vinci ;
...
site:Sede_2217
a cis:Site, poiapit:PointOfInterest ;
rdfs:label "Museo Nazionale ... Leonardo da Vinci" ;
cis:siteAddress
address:Indirizzo_della_Sede_Museo_scienza_Leonardo_da_Vinci ;
civapit:hasGeometry geometry:geometry_Museo_Leonardo_da_Vinci .
...
contactPoint:Contatti_Museo_Leonardo_da_Vinci
a smapit:OnlineContactPoint ;
smapit:hasEmail email:email_museo_Leonardo_da_Vinci ;
smapit:hasTelephoneNumber phone:phone_museo_Leonardo_da_Vinci ;
smapit:hasTelephoneNumber fax:fax_museo_Leonardo_da_Vinci ;
smapit:hasWebSite website:web_museo_Leonardo_da_Vinci .
    
```

Figure 3. Excerpt from the Lombard Museum of Science and Technology converted in RDF Turtle.

```

museum:Museo_Q3867651_Museo_Vincenzo_Vela
a cis:CulturalInstituteOrSite, cis:Museum ;
rdfs:label "Museo Vincenzo Vela" ;
cis:institutionalName "Museo Vincenzo Vela" ;
cis:hasSite site:Sede_Q3867651 ;
smapit:hasOnlineContactPoint
contactPoint:Contatti_Museo_Vincenzo_Vela .
...
site:Sede_Q3867651
a cis:Site, poiapit:PointOfInterest ;
rdfs:label "Museo Vincenzo Vela" ;
cis:siteAddress address:Indirizzo_della_Sede_Museo_Vincenzo_Vela ;
civapit:hasGeometry geometry:geometry_Museo_Vincenzo_Vela .
...
contactPoint:Contatti_Museo_Vincenzo_Vela
a smapit:OnlineContactPoint ;
smapit:hasEmail email:email_museo_Vincenzo_Vela ;
smapit:hasTelephoneNumber phone:phone_museo_Vincenzo_Vela ;
smapit:hasWebSite website:web_museo_Vincenzo_Vela .
    
```

Figure 4. Excerpt from The Swiss Vela museum converted in RFD Turtle.

It should be noted that the two museums, initially formalised in different ways on their original portal, are finally described in a common interoperable RDF format. This translation process leads, in this case, to information loss because there is not a full match between the initial format and the ontological one. The ontology is not expressive enough to represent all fields of the original data sources, although it contains more classes and properties than the original file format. For instance, the *number of visitors* is not included in the Cultural-ON ontology.

In the conversion process, the mapping from the initial input data format to the final RDF format would need to be configured for each data source. This requires knowing the OWL syntax and understanding the classes and properties of the selected ontologies.

To simplify the conversion process, an internal vocabulary was created to describe in a homogeneous and simple way data coming from different sources, without knowing the details of the ontology and further separate the input from the output. The main advantage of having this vocabulary is to hide the complexity of the ontology in the mapping process. The internal vocabulary is organized in categories, that represent contexts or ontologies; each category contains classes; each class has a number of fields. For instance, to describe museums we have defined the *museum Cultural-ON* category; this category contains classes, such as *museum* and *discipline*, and fields, such as *geographical coordinates*.

Thanks to the internal vocabulary, the conversion process is divided in two steps (see Figure 5):

- the conversion from the input data format to the internal vocabulary (*input mapping*)
- the conversion from the internal vocabulary to the ontological LOD format (*output mapping*).

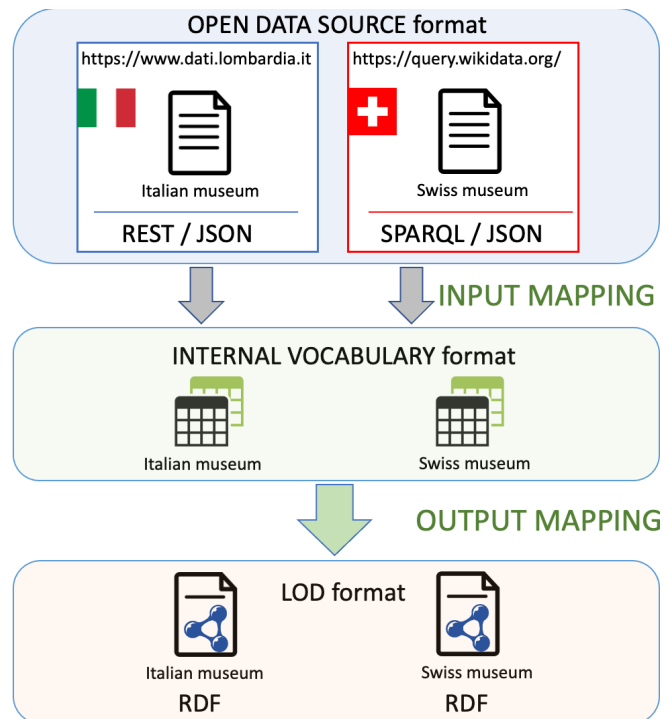


Figure 5. Two step conversion process: the museum data example.

Going back to the museum example, the two datasets, originally described in different formats and with different descriptive fields, are translated by means of the input mapping specifications into a common format, which is described by the internal vocabulary (defined in the output mapping). The resulting datasets are then converted to the LOD format, according to standard ontologies, by means of the output mapping specification. This guarantees standardization and semantic interoperability.

While it is necessary to configure the input mapping of each imported dataset towards the internal vocabulary, the output mapping of a specific category (e.g., museum) to the corresponding LOD format has to be configured only once. The first step can be accomplished by a user who knows the input format, the domain, and the internal vocabulary; the second step requires a wide knowledge of the ontologies and of the OWL language.

This mechanism that converts Open Data into Linked Open Data based on two independent and configurable steps is the peculiar feature of GIOCONDa. With respect to other LOD frameworks, we introduced a novel flexible and dynamically configurable system that simplifies the conversion and production of LOD.

It is worth noting that the conversion process may require a certain amount of time to produce LOD, which may vary from seconds to minutes according to the complexity of the data sources and the mapping specifications. This makes the LOD dataset not appropriate for real-time applications, although dataset updates can take place at regular intervals.

IV. GIOCONDA LOD PLATFORM

The GIOCONDa LOD platform [26] is mainly oriented to domain and ontology experts, who operate using their own user accounts to create and modify datasets. Public administrations can submit new datasets for conversion into LOD. The LOD datasets are also made accessible through a public portal without registration.

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.

The web app presents a menu consisting of different items: dataset catalogue, input mapping, and output mapping.

A. LOD Datasets

The catalogue shows the list of the existing datasets, as shown in Figure 6, and enables the creation of a new LOD dataset by converting an existing open dataset on the basis of the input and output mapping configuration. The system supports dataset updates at regular intervals (e.g., for air quality measurements) and propagates the changes to the RDF representation.

Name	Description	Category	RDF file	RDF view	Update	Options	Add New Dataset
TI ostelli	TI ostelli - Elenco	Turismo	TI_ostelli.rdf Downloaded: 40	Classes	Update 17/11/2021 20:56	Visible	
TI musei	TI musei - Musei del	Cultura	TI_musei.rdf Downloaded: 53	Classes	Update 16/11/2021 12:31	Visible	
TI hotel	TI Hotel - Elenco hotel	Turismo	TI_hotel.rdf Downloaded: 63	Classes	Update 04/10/2021 18:47	Visible	
TI capanne alpine	TI Capanne alpine -	Turismo	TI_capanne_alpine.rdf Downloaded: 55	Classes	Update 16/11/2021 23:32	Visible	
TI campeggi	TI Campeggi -	Turismo	TI_campeggi.rdf Downloaded: 53	Classes	Update 16/11/2021 23:05	Visible	
TI B&B	TI B&B - Elenco B&B	Turismo	TI_B&B.rdf Downloaded: 66	Classes	Update 04/10/2021 19:47	Visible	

Figure 6. LOD datasets catalogue.

By clicking on the “map view” button, it is possible to visualize data on the map, whenever they have geographical coordinates, as shown in Figure 7.

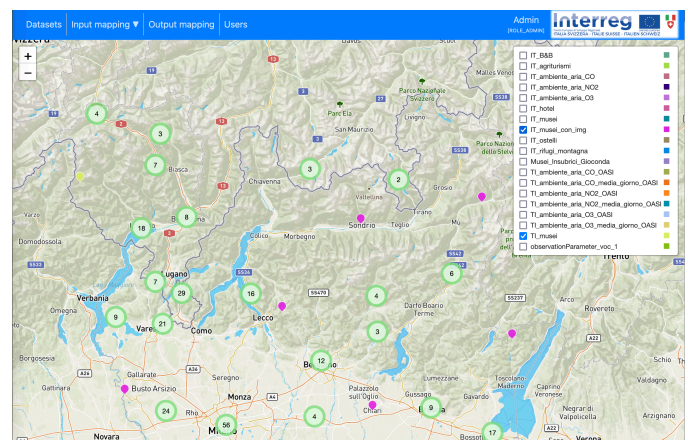


Figure 7. LOD Datasets visualized in the map.

Approximately 25 datasets about accommodation, museums, and air quality were boosted to LOD through the GIOCONDa platform. Concerning validation, the output mapping process guarantees by design and implementation that the produced datasets are accurately serialized in RDF format conforming to the selected ontologies. A further manual checking was accomplished on some resources of each typology.

From the LOD datasets page (shown in Figure 6), it is possible to add a new dataset by clicking on the corresponding button. This action activates the conversion process, driven by the input and output mapping on the specific data sources. The “Add new dataset” function was used to generate the single datasets shown in Figure 6, but it could also be used to combine more sources to produce a unified dataset (see Figure 8). For instance, to create the “Insubric museum” dataset, it is possible to select the Italian and Swiss museums as data sources, “IT musei” and “TI musei” respectively.

Figure 8. New dataset creation.

B. Input Mapping

The Input mapping concerns the configuration of the conversion from the input format to the internal vocabulary. Together with the output mapping, it enables to configure the conversion from different input data sources to the final LOD format.

The system accepts input data retrieved from sources in different formats, such as JSON, CSV, and XML, and using different import modes, such as Rest APIs, SOAP APIs, and SPARQL Queries. For each input format and import mode the conversion towards the internal vocabulary is configured through the input mapping.

Figure 9 clarifies how the mapping mechanism works: the first column shows the fields of the original data source; the second and third columns concern the internal vocabulary, where in particular the second identifies the category, and the third the field.

Categories and fields of the internal vocabulary are predetermined and selected from a drop-down menu, while the source fields must be entered by hand, according to the data structure in use, as explained later.

In the example shown in Figure 9, fields of the “Cultural-ON museums” category are used; the “email_sede” field of the initial source, representing the email of the museum site, for example, is mapped into the “email” field of the “museum Cultural-ON” category of the internal vocabulary. In some cases, it is possible to group multiple fields of the data source into a single field of the internal vocabulary; for example, to compose the field “full_address”, more fields of the input source are used.

In this mapping, particular attention is dedicated to how geospatial data are represented [27]. This is essential to guarantee interoperability and efficient sharing of information across different regions and national standards. The Cultural-ON ontology assumes, by default, that spatial data are represented in the geocentric Datum WGS84 and that the coordinates are expressed in terms of latitude and longitude. Therefore, data are transformed into this system when they are

imported into the GIOCONDA platform and appropriate metadata are added to make the Coordinate Reference System (CRS) explicit.

Figure 9. Input mapping.

The input mapping mechanism is the same for the different import modes, which only differ for the way the original data structure is imported.

For instance, Figure 10 shows an example of configuration of a JSON data source imported through Rest APIs.

Figure 10. Rest API configuration.

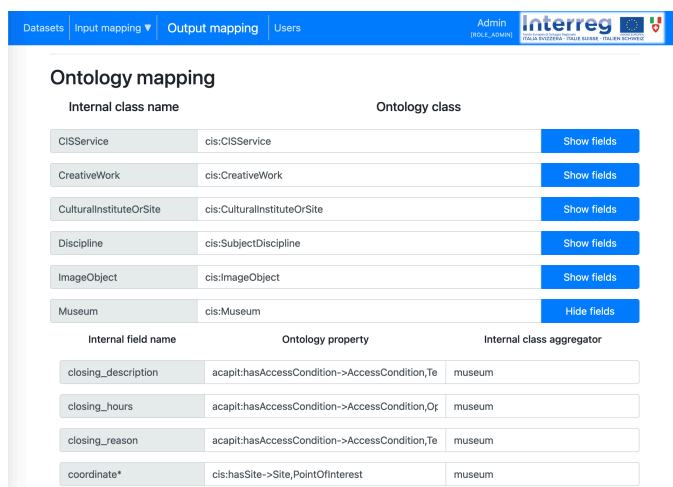
On this page the data source is first defined through a name, a URL, and the licence, followed by a testing section and the input mapping section as presented in Figure 9. In the “Source testing” section there is a “run request” button, which is used to verify the proper functioning of the specified REST call; this button visualizes all the returned fields of the input source together with their values. This call is also useful to examine the field names to be used as input sources in the configuration of the input mapping.

C. Output Mapping

From the output mapping page, it is possible to create, modify and extend the internal vocabulary, and define its mapping to the ontology. This process requires a deep knowledge of ontological concepts and existing reference ontologies. Nevertheless, this mapping has to be done only once for each category by an expert.

As already said, the internal vocabulary consists of several categories, similar to contexts or ontologies; each category contains classes, with a number of associated fields. Examples of categories include museums, addresses, accommodations, etc.

As shown in Figure 11, the output mapping defines the match between internal vocabulary classes and ontology classes, and between fields of the internal vocabulary and object and datatype properties of the ontology; this is visible by activating the “Show fields” button.



Internal class name	Ontology class	
CISService	cis:CISService	Show fields
CreativeWork	cis:CreativeWork	Show fields
CulturalInstituteOrSite	cis:CulturalInstituteOrSite	Show fields
Discipline	cis:SubjectDiscipline	Show fields
ImageObject	cis:ImageObject	Show fields
Museum	cis:Museum	Hide fields

Internal field name	Ontology property	Internal class aggregator
closing_description	acapit:hasAccessCondition->AccessCondition,Te	museum
closing_hours	acapit:hasAccessCondition->AccessCondition,Or	museum
closing_reason	acapit:hasAccessCondition->AccessCondition,Te	museum
coordinate*	cis:hasSite->Site,PointOfInterest	museum

Figure 11. Output mapping: class and field match.

It is worth noting that only categories and fields of the internal vocabulary defined in the output mapping can be used in the input mapping (but not classes), providing in this way a simplified version of the data structure for non-expert users.

Another action that takes place in the output mapping is the interlinking configuration. For this purpose, a specific interlinking module was developed and integrated into the GIOConDa LOD platform. The next section describes how the interlinking process works.

V. THE INTERLINKING MODULE

In order to boost a dataset to 5* level, it is necessary to enrich the RDF file with connections towards external sources. The rules to create cross-reference links towards external datasets, such as Wikidata, are defined in specific files, generated using the Silk Link Specification Language, serialized as .xml using the Linkage Rule Editor.

Once the Silk files are generated, it is necessary to configure the interlinking and then activate it.

The configuration is accomplished in the output mapping page, where one or more interlinking files can be associated to

each category of the internal vocabulary (e.g., museum). This association is called “interlinking configuration”.

The activation takes place on the datasets page, where it is possible to enable or disable interlinking on a specific dataset.

The rest of this section describes the Linkage Rule Editor, the interlinking configuration, and the interlinking activation in more detail.

A. The Linkage Rule Editor

The Linkage Rule Editor is part of the Silk Linked Data Integration Framework [18]. This open-source framework for integrating heterogeneous data sources was selected for the easiness to be adapted and integrated into the GIOConDa LOD platform. In general, the primary use cases of Silk include: generating links between different data sources and applying transformations to data from structured data sources; Linked Data editors can use Silk to set up RDF links from their data sources to other data sources on the web. Silk is powerfully based on the Linked Data paradigm, where, on the one hand, RDF provides an expressive data model for the representation of structured information; on the other hand, RDF links are set up between entities in different data sources.

The rules to create cross-reference links towards external datasets are defined in the Linkage Rule Editor page (Figure 12).

Link specifications can be created using the graphical interface or manually in XML. Using the Silk Link Specification Language (Silk-LSL) declarative language, developers can specify which types of RDF links must be detected between data sources, as well as the conditions that data elements must meet to be interconnected. These link conditions can combine various similarity metrics and can take into account the graph around a data element (entity), which is identified using the RDF path language.

Silk accesses the data sources via the SPARQL query language, supporting the use of local and remote SPARQL endpoints.

Figure 12 shows the creation of the owl:sameAs link between different entities based on the comparison of their geo-location coordinates. In particular, the first sourcePath and targetPath take the latitudes, while the latter two take the longitudes both from instances of a museum of the GIOConDa LOD dataset and of Wikidata. We use the following criteria to match entities from the two data sources: if two entities of the same type (e.g., one museum from the GIOConDa LOD dataset, another from Wikidata) are located within 10 meters each other, then they are considered as the same entity.

The process corresponding to the rules defined in the example of Figure 12 is the following. First, the coordinates are obtained via the corresponding SPARQL queries; then a numerical comparison (*numericEquality*) is performed both between the latitudes and between the longitudes, with a precision up to the 4th decimal place and a maximum threshold of 0.0001. This is equivalent to about ten meters in the physical world. If both pass the equality comparison (*min* represents a logical “and” aggregation), the interconnecting link is created by declaring an owl:sameAs relationship. In simple words, two museums are considered the same if their distance is below ten meters.

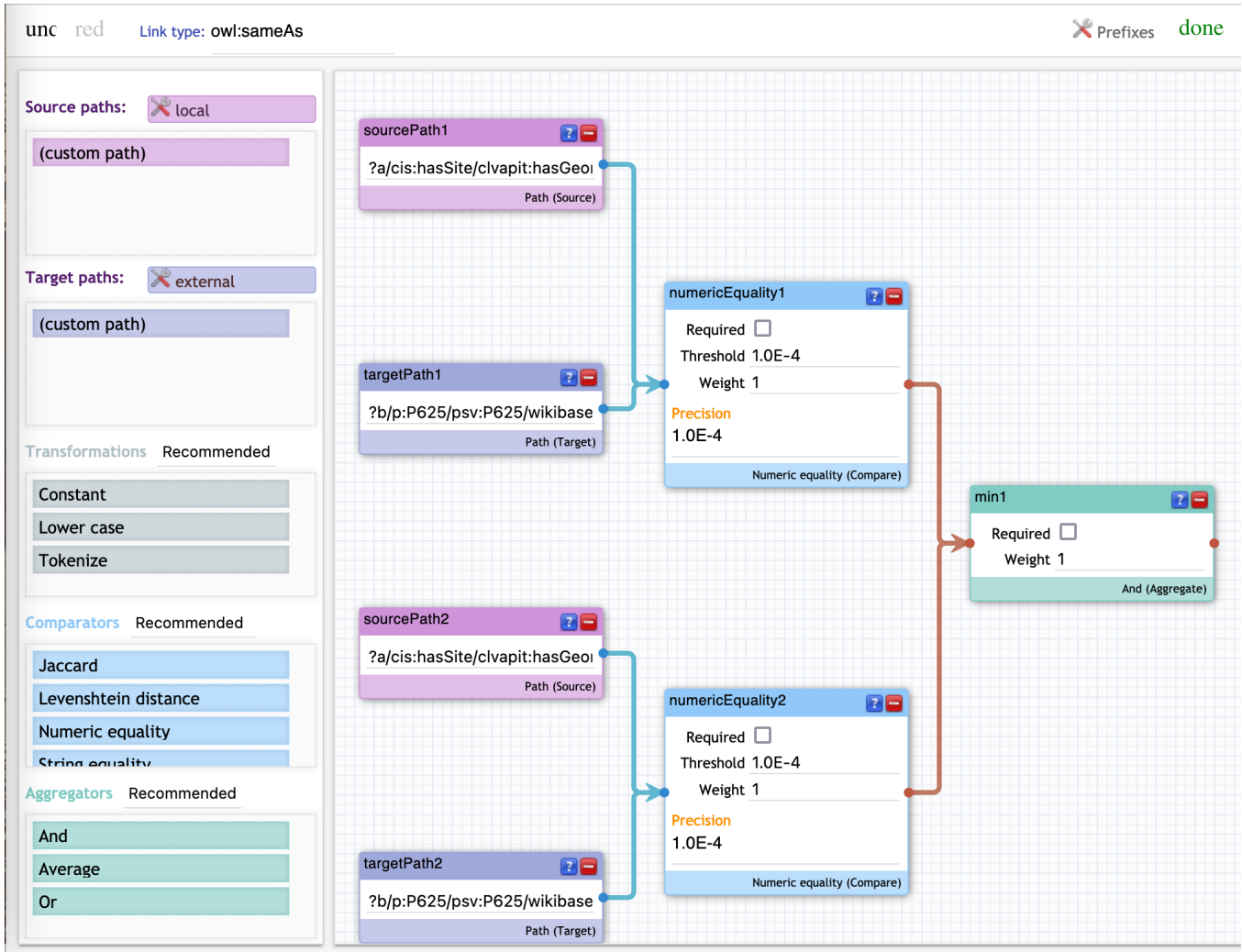


Figure 12. Example of the Linkage Rule Editor use.

B. Interlinking configuration

Figure 13 shows an example of the interlinking configuration within the Output mapping page for a specific category. On this page it is possible to upload a rule set from an external file, remove an existing set, create a new rule set, or modify an existing one, by opening the Linkage Rule Editor page (see also Figure 12).

It is worth noting that the Linkage Rule editor is a page fully integrated into the GIOCOOnDa LOD platform, after adaptation of the Silk original editor page.

C. Interlinking activation

The interlinking activation takes place on the configuration page of a dataset, where, through selection, it is possible to activate or deactivate one or more configuration files on the current dataset (Figure 14).

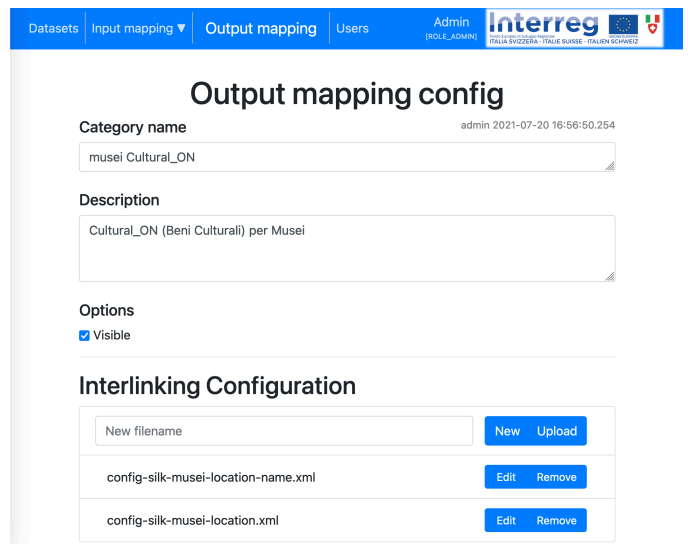


Figure 13. Example of interlinking configuration for a category in the output mapping configuration page.

The ability to simultaneously use multiple configurations allows interlinking between the generated RDF dataset and multiple data sources to be performed; for example, interlinking with the Wikidata endpoint with one configuration file, while interlinking with Open Street Map with a second one.



Figure 14. Example of interlinking activation.

VI. LOD PUBLISHING PROCESS: USE CASES

To better understand how the process of publishing a new dataset works, three different use cases can be distinguished:

- request for publication of a dataset with the **same structure** of an existing one (for example, a new dataset structured as the Lombard museums). Since the mapping from the input format to the internal vocabulary of that category is already configured, the conversion will simply take place by duplicating the existing input mapping configuration; the output mapping is already configured;
- request for publication of a dataset with a **new initial structure that can be mapped** into the existing internal vocabulary (for example, a new dataset about museums, structured in a different way compared to the existing datasets). In this case it will be necessary to configure the input mapping from the input format to the internal vocabulary of that category; the output mapping is already configured;
- request for publication of a dataset with a **new initial structure that cannot be mapped** into the existing internal vocabulary (for example, a new dataset relating to bike sharing, for which a vocabulary has not been defined yet). In this case it will be necessary: to look for an ontology that models the domain; to extend the internal vocabulary by adding the category “bike sharing”; to define the correspondence between the internal vocabulary and the domain ontology (output mapping); to configure the matching between the initial format and internal vocabulary (input mapping).

In summary, in the first case, there is no need to configure input and output mapping, since they already exist; in the second case, a new input mapping configuration is required; in the third one both the input and the output mapping need to be configured.

In all cases, once the two configurations of input and output mapping have been defined, it is possible to proceed with the creation of the new dataset.

VII. CONSUMING GIOCONDA OPEN DATA

Linked Open Data is useless if it cannot be extracted in a format convenient for processing and exploiting the enrichment and interlinking of LOD.

To this end, the GIOCONda platform provides a SPARQL endpoint that can be queried to extract LOD data from the RDF data sources stored in the Virtuoso server, providing ontology-based data access. The endpoint is available at the URL: <https://gioconda.supsi.ch:8890/sparql>. Thanks to this endpoint, anyone can fetch homogeneous data to implement new applications that can exploit the enrichment and interlinking of LOD, with a limited knowledge of the complex notions of ontologies and RDF.

As case studies, two applications that provide new visualizations of the data are illustrated. The first provides an overview of the museums in Tessin and Lombardy, whose data have been already described in the previous section. The second application provides a list of holiday accommodations in Tessin and Lombardy.

A. Overview of museums in Tessin and Lombardy

As first example, we implemented an application that provides an overview of museums in Tessin and Lombardy. The application is interactive and allows zooming and focusing on a particular museum to access its main information: name of the museum, type of museum, geographical location, address, and a picture. These attributes are collected from public sources based on Linked Open Data (namely: Regione Lombardia portal and Wikidata), converted and interlinked in the GIOCONda platform.

The purpose of the application is to allow people to explore and have an overview of the distribution of various types of museums in the two regions. People who have to take important decisions related to the presence of museums in a certain area, such as opening a new museum or establishing a new partnership among different museums covering the same subjects, may use this application.

To implement the application, the visual analytics tool Tableau [28] was used. Tableau allows the integration of a large number of data sources, but unfortunately it does not allow the direct integration of SPARQL endpoints. To overcome this problem, the data.world platform [29] was used as a bridge to connect the GIOCONda SPARQL endpoint with the Tableau platform. Data.world is a cloud-based service that allows the creation of a repository of data that can be integrated with dozens of applications for analytics and visualization. In that way, we can produce visual representations of data accessing data directly from the GIOCONda platform (see Figure 15).

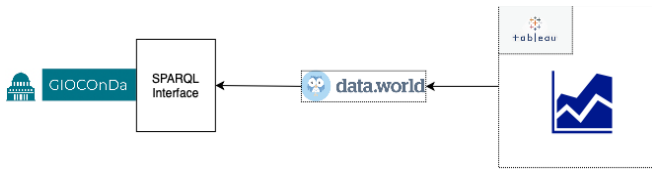


Figure 15. Structure of integration of GIOCOOnDa, Data.world, and Tableau.

With this architecture, the dashboard illustrated in Figure 16 was implemented in Tableau. This dashboard consists of 4 coordinated views. On the top left there is a bar chart that shows the quantity of the different types of museums in the two regions.

The user can filter out specific elements to see only data related to a particular region (Lombardy/Tessin) or location (city). A detailed view on the right provides a picture of the museum (if existing) and its details (name, location, type, and number of visitors per year). On the bottom left, the largest area is dedicated to a dot representation of museums on a map, and a tabular list of them on the right. Hovering the cursor of the mouse over a particular dot opens a new window with the details of the museum. Using this dashboard, the viewer can 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.

B. Overview of accommodations in Tessin and Lombardy

The second case study provides a list of holiday accommodations in Tessin and Lombardy. Like in the previous application, the purpose is to explore the different types of

accommodations across the two regions in a uniform way. For this new application the source data are taken from Wikidata (that provides a list of Italian accommodations) and Ticino Turismo (for the Swiss data). GIOCOOnDa regularly fetches data from these data sources, in order to keep data regularly updated, and store these open data in its internal data structures.

To extract the data from the GIOCOOnDa platform, the same infrastructure described in the previous case study was used. In this case, new SPARQL queries had to be written and a new visual interface to extract and visualize data relevant to this new goal was implemented.

Figure 17 shows the dashboard implemented for this purpose. As in the previous application, we kept the approach of using a dot map to represent locations of accommodations. The circle colours encode the different types of accommodations (hotel, hostel, B&B, ...). A bar chart at the top shows the percentages of each type of accommodation in the two regions. A possible scenario for usage of this visualization is the following. An entrepreneur wants to invest in opening a new hospitality structure in Ticino. He/she may check this visualization and see that Agriturismo (farmhouses) and B&B are largely underrepresented in Tessin with respect to Lombardy, so he/she decides to open one or more structures of this typology. This could be an example of a decision driven by data and supported by this application.

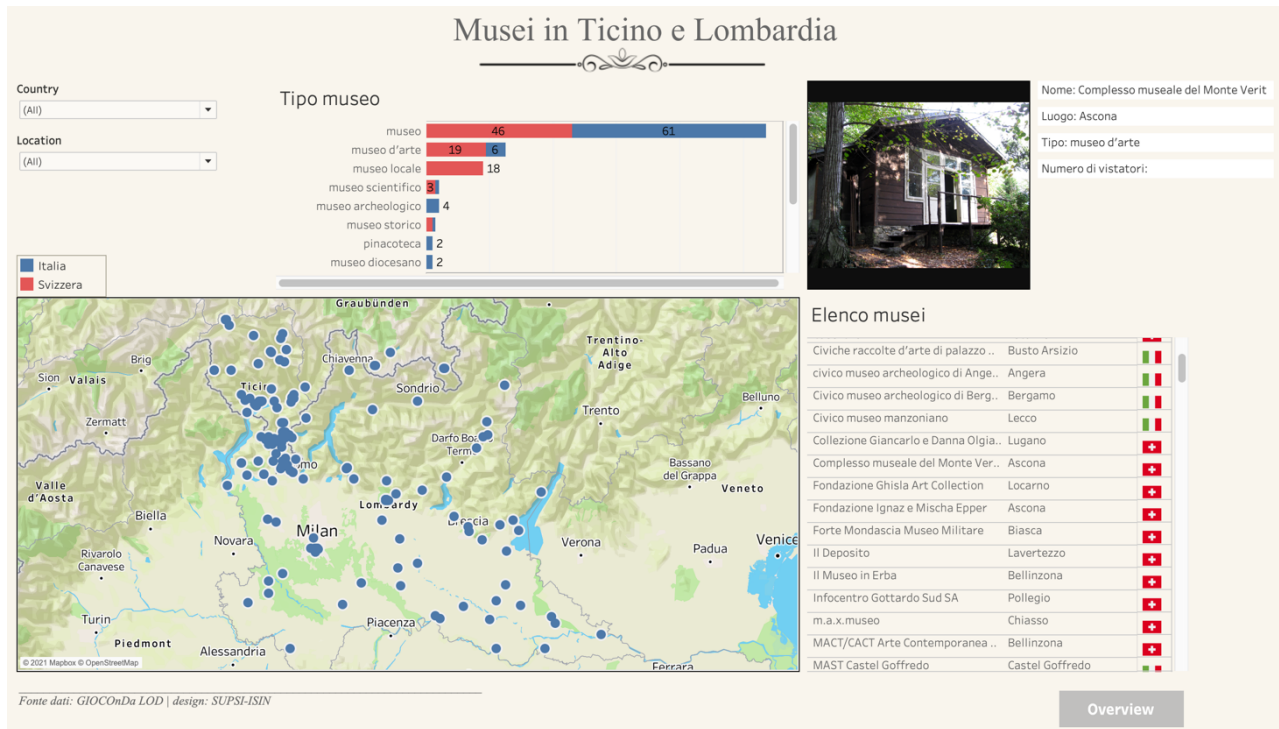


Figure 16. Overview of museums in Ticino and Lombardy.

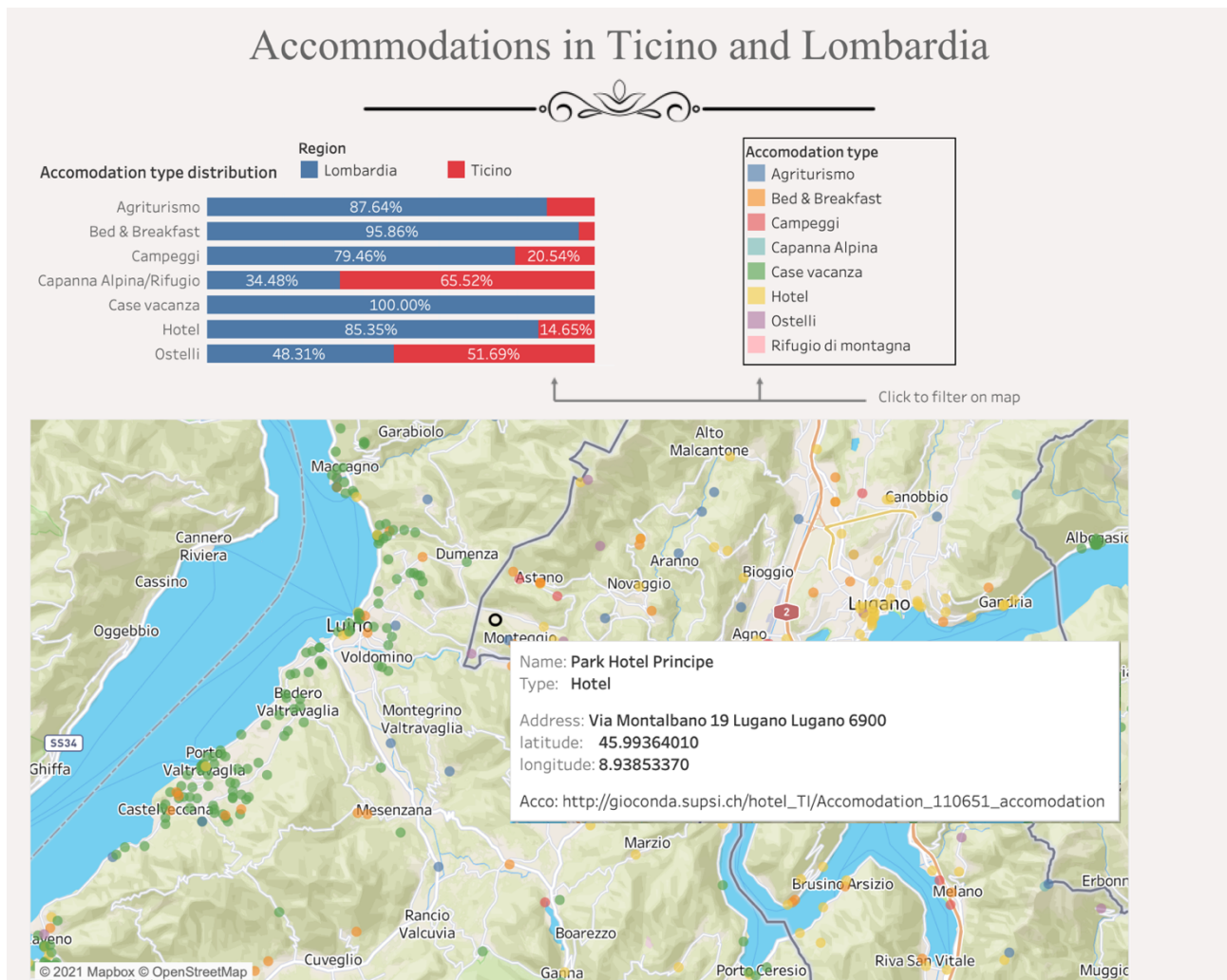


Figure 17. Overview of holiday accommodations in Ticino and Lombardy.

VIII. CONCLUSION AND FUTURE WORK

This paper has presented a platform that facilitates the process of conversion of open data to Linked Open Data and some examples of applications that consume LOD to reveal information that cannot be captured in a different way.

The GIOCOOnDa LOD platform contains a number of conversion configurations that are already available to translate different data sources to LOD in various domains.

The complexity of mapping existing data to standard ontologies is one of the major issues preventing a larger diffusion of LOD. The GIOCOOnDa platform reduces the complexity of this process that would require deep knowledge of ontologies and programming skills.

If we analyse the conversion process in further detail, it is possible to distinguish different use cases and complexity levels, according to the initial structure of the dataset to be converted.

If the dataset has the same structure of an existing one (for example, a new dataset structured as the Lombard

museums), the conversion is very simple, since the input mapping is similar to an existing one (it just needs to be duplicated) and the output mapping is already defined. However, the platform is also flexible and extensible, and enables to import and convert other datasets: for example, the conversion to LOD of a new dataset about museums, with a structure that can be mappable into the existing internal vocabulary, only requires the configuration of the input mapping from the initial format to the internal vocabulary, because the output mapping is already configured. More labour-intensive but still possible is to convert a dataset with a new structure, not mappable into the existing internal vocabulary; for example, a new dataset about bike sharing.

In addition to the conversion of structured open data to standard RDF open data, another important step of the adopted methodology to produce LOD is the identification and creation of interlinks between datasets. A specific interlinking module was developed to configure and activate the process of identification of cross-reference links towards external datasets. The integration of the interlinking module

in the GIOCOOnDa LOD platform enables lifting datasets to 5* level, creating added value through an Extract-Transform-Load (ETL) pipeline. This is demonstrated, for instance, in a showcase that presents data about museums of the Insubric region taken both from the GIOCOOnDa LOD datasets and from Wikidata.

Despite the benefits offered by the platform to publish LOD datasets (visual interface, configurability, extensibility), it also presents some limitations: the main one is the possibility of information loss during the conversion to LOD if there are fields not represented in the selected ontology. A possible solution would be the extension of the selected ontologies with additional fields and the publication of the new version with appropriate documentation.

By using the platform functionalities, a consistent number of datasets were created and will be produced in the future. However, it can be difficult to interpret data in a tabular way; data gain more value when they are visualized. By means of visualization tools it is possible to communicate data findings and identify critical information to pull insights.

To demonstrate the potential of Linked Open Data some applications were developed concerning museums and accommodations. In both cases the applications provide an overview of the distributions of these entities in the Insubric area, supporting decision making in these domains. Once the number of datasets increase, new opportunities will be available to develop consuming applications.

In conclusion, it is possible to state that the GIOCOOnDa project has reached its objectives thanks to the LOD platform and the connected applications: a consistent number of LOD datasets were published and examples of data consuming applications were implemented to show the LOD potential.

In the future the platform will be further developed, completing and improving some parts, such as increasing the internal vocabulary to publish new LOD datasets, simplifying the user interface and refining the interlinking rule definition.

ACKNOWLEDGMENT

We acknowledge the European Regional Development Fund, the Tessin Canton and the Swiss Confederation for the financial support provided to the project. We also acknowledge all the partners involved in the project for their contribution to the different project activities.

REFERENCES

[1] L. Sommaruga, N. Catenazzi, D. Bertacco, and R. Mazza, "From Open Data to Linked Open Data - The GIOCOOnDa LOD platform", ALLDATA 2021 (The Seventh International Conference on Big Data, Small Data, Linked Data and Open Data), April 18, 2021 to April 22, 2021 - Porto, Portugal.

[2] GIOCOOnDa, Integrated and holistic management of the Open Data life cycle (Gestione integrata e olistica del ciclo di vita degli Open Data) [italiasvizzera.eu/it/b/78/gestioneintegrataeolisticadelciclodivitadegliopendata \[retrieved: 2021.11.30\].

\[3\] Open Data, Regione Lombardia, <https://dati.lombardia.it> \[retrieved: 2021.11.30\].

\[4\] Arpa Lombardia, <https://www.arpalombardia.it/> \[retrieved: 2021.11.30\].

\[5\] Ticino Turismo, <https://www.ticino.ch> \[retrieved: 2021.11.30\].

\[6\] OASI, Osservatorio Ambientale della Svizzera italiana, <https://www.oasi.ti.ch> \[retrieved: 2021.11.30\].

\[7\] T. Berners-Lee, "Linked Data", <http://www.w3.org/DesignIssues/LinkedData.html>, 2009 \[retrieved: 2021.11.30\].

\[8\] T. Heath and C. Bizer, *Linked Data, "Evolving the Web into a Global Data Space"*, Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool Publishers, 2011.

\[9\] The Lucero Project, <https://lucero-project.kmi.open.ac.uk/> \[retrieved: 2021.11.30\].

\[10\] L. Ding, T. Lebo, J. S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. G. Zheng, Z. Shangquan, J. Flores, D. L. McGuinness, and J. A. Hendler, "TWC LOGD: A portal for linked open government data ecosystems", *Journal of Web Semantics*, vol. 9\(3\), pp. 325-333, Sep. 2011, doi: 10.1016/j.websem.2011.06.002.

\[11\] Open Data e Linked Data, <https://www.beniculturali.it/open-data-e-linked-data> \[retrieved: 2021.11.30\].

\[12\] A. Meherhera, I. Mekideche, L. Zemmouchi-Ghomari, and A. Réda Ghomari, "A Survey of Current Approaches for Transforming Open Data to Linked Data". 4th Edition of the National Study Days on Research on Computer Sciences, JERI'2020, Jun 2020, Saida, Algeria. hal-03211592, <https://hal.archives-ouvertes.fr/hal-03211592/document> \[retrieved: 2021.11.30\].

\[13\] F. Bauer and 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, <https://www.recep.org/LOD-the-Essentials.pdf> \[retrieved: 2021.11.30\].

\[14\] W3C, "Best Practices for Publishing Linked Data", W3C Working Group Note 09 January 2014, <https://www.w3.org/TR/ld-bp/> \[retrieved: 2021.11.30\].

\[15\] Agenzia per Italia Digitale, Guidelines for semantic interoperability through Linked Open Data \("Linee Guida per Interoperabilità semantica attraverso i Linked Open Data"\), \[http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf\]\(http://www.agid.gov.it/sites/default/files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf\) 2012 \[retrieved: 2021.11.30\].

\[16\] Ontologie e Vocabolari Controllati, <https://github.com/italia/daf-ontologie-vocabolari-controllati> \[retrieved: 2021.11.30\].

\[17\] G. Lodi, "OntoPiA - The network of ontologies and controlled vocabularies for public administration \(OntoPiA - La rete di ontologie e vocabolari controllati per la pubblica amministrazione\)", *Open Data Sicilia - raduno annuale*, 9/10 novembre 2018, <http://ods2018.opendatasicilia.it/presentazioni/Lodi-OntoPiA.pdf>, 2018 \[retrieved: 2021.11.30\].

\[18\] Silk, *The Linked Data Integration Framework*. <http://siliframe.org/> \[retrieved: 2021.11.30\].

\[19\] Open Link Virtuoso, <https://virtuoso.openlinksw.com> \[retrieved: 2021.11.30\].

\[20\] C. Bizer and R. Cyganiak, "D2R Server - Publishing Relational Databases on the Semantic Web", 5th International Semantic Web Conference \(ISWC 2006\), Athens, USA,](https://progetti.interreg-</p>
</div>
<div data-bbox=)

- November 2006, <http://richard.cyaniak.de/2008/papers/d2r-server-iswc2006.pdf>, 2006 [retrieved: 2021.11.30].
- [21] R2ML: RDB to RDF Mapping Language <https://www.w3.org/TR/r2rml> [retrieved: 2021.11.30].
- [22] Musei riconosciuti da Regione Lombardia <https://www.dati.lombardia.it/Cultura/Musei-riconosciuti-da-Regione-Lombardia/3syc-54zf> [retrieved: 2021.11.30].
- [23] Museo Vela <https://www.wikidata.org/wiki/Q3867651> [retrieved: 2021.11.30].
- [24] Cultural-ON (Cultural ONtology): Cultural Institute/Site and Cultural Event Ontology, <https://w3id.org/italia/onto/Cultural-ON> [retrieved: 2021.11.30].
- [25] Social Media / Contact and Internet ontology - Italian Application Profile <https://w3id.org/italia/onto/SM> [retrieved: 2021.11.30].
- [26] GIOCONDa LOD platform <https://gioconda.supsi.ch/> [retrieved: 2021.11.30].
- [27] J. F. Toro Herrera, D. Carrion, M. A. Brovelli, N. Catenazzi, L. Sommaruga, and D. Bertacco “Geospatial Data dissemination in the GIOCONDA project”, extended abstract, ASITA (Federazione delle Associazioni Scientifiche per le Informazioni Territoriali e Ambientali, <http://atti.asita.it/ASITA2021/Pdf/039.pdf>) July 2021, [retrieved: 2021.11.30].
- [28] Tableau: Business Intelligence and Analytics Software <https://www.tableau.com/>, [retrieved: 2021.11.30].
- [29] “The catalog for metadata and data management”, <https://data.world>, [retrieved: 2021.11.30].