From Open Data to Linked Open Data
The GIOCOOnDa LOD platform

Authors: L. Sommaruga, N. Catenazzi, D. Bertacco, R. Mazza

Lorenzo Sommaruga
Department of Innovative Technologies
University of Applied Sciences and Arts of Southern Switzerland (SUPSI)
CH-6962 Lugano, Switzerland

e-mail: lorenzo.sommaruga@supsi.ch

ALLDATA 2021, April 18 - 22, 2021 - Porto, Portugal
Presenter: Lorenzo Sommaruga

- Graduated in Information Science at the University of Milan, PhD in Psychology at Nottingham University (UK)
- he worked at different universities in Europe
- he is currently professor at the Dept. of Innovative Technologies of the University of Applied Sciences and Arts of Southern Switzerland (SUPSI) and lecturer of web technologies
- researcher in the field of web applications, e-learning and open data in European projects; responsible for the Linked Open Data and Educational Technologies research area
Presenter’s research group

*Educational technologies and open data* research area, SUPSI DTI ISIN


- research units:
  - Educational Technologies
    - Learning Management Systems
    - Tools for interactive content and interactive video
    - Open Educational Resources and open badges
    - Learning analytics
  - Linked Open Data
    - Semantic web and linked data technologies
    - Data modeling through ontologies
    - Open Data license
Presentation Content

• Context: the GIOCOOnDa project
• Target groups and data sources
• Methodology to Publish Linked Open Data
• The Conversion Process from Open Data to LOD
• The GIOCOOnDa LOD Platform
• Conclusions
Context: the GIOCOOnDa project

• 24 month project, funded by the Interreg V-A Italy-Switzerland Programme (2019-2021)

• It aims to create value by developing information products based on the re-use of public Open Data

• It involves the creation of the GIOCOOnDa LOD platform for the publication of Linked Open Data
  – the platform allows conversion of existing 3* Open Data to 5* Open Data, according to the well-known 5-star deployment scheme
Target groups

• The platform is mainly oriented to domain and ontology experts who, once authenticated, can create and modify datasets.

• A general public portal is also available, where the datasets produced in the LOD platform are made accessible and searchable.

• Public administrations can submit new datasets for conversion into LOD.
Data sources

• The project focuses on data from the Insubric region, a cross-border territory across Italy and Switzerland.

• about museums, accommodation facilities and environment
Data sources

**REGIONE LOMBARDIA**
- Farmhouses
- B&Bs
- Camping
- Holiday apartments
- Hotels
- Museum
- Hostels
- Alpine huts

**ARPA**
- Environmental data (CO, NO₂, O₃)

**TICINO TURISMO**
- Farmhouses
- B&Bs
- Camping
- Hotels
- Hostel
- Alpine huts

**OASI - Cantone Ticino**
- Environmental data (CO, NO₂, O₃)

**Wikidata**
- Museums
## Methodology for LOD publishing

<table>
<thead>
<tr>
<th>Methodological STEPS</th>
<th>Implementation in GIOCOOnDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>selection of dataset</td>
<td>resulted from the need analysis phase</td>
</tr>
<tr>
<td>data cleaning</td>
<td>assumption that the selected datasets are already published as “clean” open data</td>
</tr>
<tr>
<td>analysis, ontology selection, RDF modelling</td>
<td>ontologies selected from the OntoPia network: Cultural-ON for museums, ACCO for accommodations ...</td>
</tr>
<tr>
<td>enrichment</td>
<td>datasets enriched with metadata (DCAT-AP)</td>
</tr>
<tr>
<td>interlinking</td>
<td>interlinks to other datasets (e.g. Wikidata) using Silk</td>
</tr>
<tr>
<td>validation</td>
<td>guaranteed by design and manual checking</td>
</tr>
<tr>
<td>publication</td>
<td>using Openlink Virtuoso Universal Server queries through SPARQL endpoint</td>
</tr>
</tbody>
</table>
The Conversion Process from Open Data to LOD

• It is a **complex process** that depends on the initial data format and the final ontological format

• the **most frequently adopted approach** is the implementation of ad-hoc middleware that requires ontological and programming skills (e.g., D2R or R2RML)

• in the GIOCOOnDa LOD platform, the **complexity** of the conversion is **simplified** by defining a converter that an expert can use to configure the conversion

• This process is explained through a simple **example**: we would like to convert two different datasets about museums into a common interoperable format
Example: initial format IT museum

Source: https://www.dati.lombardia.it/Cultura/Musei-riconosciuti-da-Regione-Lombardia/3syc-54zf
Example: initial format CH museum

Source: https://www.wikidata.org/wiki/Q56876232
Example: final RDF format

IT and CH museums

Selection of the Cultural-On ontology to model the domain

Search for a match between the descriptive fields of the IT and CH museums and the ontology classes and properties

Final RDF interoperable format
Internal vocabulary

• **Created to simplify** the complexity of the conversion process that requires a deep knowledge of the OWL syntax
  – with the **objective** to describe in a simple way data coming from different sources

• The internal vocabulary is organized in **categories**, that represent contexts or ontologies
  – each category contains **classes**
  – each class has a number of **fields**

• **Example**: to describe museums we have defined the *Museum* category that contains **classes**, such as *museum* and *discipline*, and **fields**, such as geographical *coordinates*
Conversion process in two phases

Two steps:
1. **input mapping**: conversion from the input data format to the internal vocabulary
2. **output mapping**: conversion from the internal vocabulary to the ontological LOD format

While it is necessary to configure the input mapping of each imported dataset, the output mapping of a specific category to the corresponding LOD format has to be configured only once.
GIOCOOnDa LOD Platform

- implemented as a **Java based web application**
- it provides functionalities that enable the publication of LOD datasets and their visualization in a catalogue or in a map
Gioconda LOD architecture

Italian providers

Swiss providers

GIOCONDA LOD Platform

INPUT MAPPING

OUTPUT MAPPING

standard ontologies

GIOCONDA datl.lombardia.it

RDF data

LOD form

Users

SPARQL endpoint

VIRTUOSO

Italian providers

Swiss providers

JSON, .CSV...

INPUT MAPPING

OUTPUT MAPPING

standard ontologies

JSON, .CSV...

Users

SPARQL endpoint

VIRTUOSO
LOD Datasets

It shows the **dataset catalogue** and enables the creation of a new dataset on the basis of the input and output mapping configuration.
Map view

The GIOCOOnDa LOD Platform functionalities

LOD Datasets visualized on the map
It concerns the configuration of the conversion **from the input format** to the **internal vocabulary**

The system accepts input data

- from different sources and formats (JSON, CSV and XML)
- using different services (Rest APIs, SOAP APIs and SPARQL queries)

<table>
<thead>
<tr>
<th>Source field</th>
<th>Category</th>
<th>Internal Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>[*] tipo_chiusura</td>
<td>musei Cultural.ON</td>
<td>closing_description</td>
</tr>
<tr>
<td>[*] motivazione_chiusura_tempo_det</td>
<td>musei Cultural.ON</td>
<td>closing_reason</td>
</tr>
<tr>
<td>EPSG4326</td>
<td>musei Cultural.ON</td>
<td>coordinate_epsg</td>
</tr>
<tr>
<td>WGS84</td>
<td>musei Cultural.ON</td>
<td>coordinate_system</td>
</tr>
<tr>
<td>[*] tipologia_museo</td>
<td>musei Cultural.ON</td>
<td>discipline</td>
</tr>
<tr>
<td>[*] email_sede</td>
<td>musei Cultural.ON</td>
<td>email</td>
</tr>
<tr>
<td>[*] provincia_sede, comune_sede, indirizzo_sede, frazione_sede, cap_sed</td>
<td>musei Cultural.ON</td>
<td>full_address</td>
</tr>
<tr>
<td>[*] location_latitude</td>
<td>musei Cultural.ON</td>
<td>lat</td>
</tr>
<tr>
<td>[*] location_longitude</td>
<td>musei Cultural.ON</td>
<td>lon</td>
</tr>
<tr>
<td>[*] denominazione_museo</td>
<td>musei Cultural.ON</td>
<td>name</td>
</tr>
</tbody>
</table>
It concerns the configuration of the conversion from internal vocabulary to the ontological format.

It enables to create, modify and extend the internal vocabulary and define its mapping to the ontology.

A specific interlinking module has also been developed and integrated in the GIOCOOnDa LOD platform.
Conclusions

• Main advantages of the GIOCOOnDa LOD platform
  – no programming skills to configure the conversion from Open Data to Linked Open Data
  – reduced complexity thanks to two step conversion process (input/output mapping)
  – flexibility and dynamic configurability
  – Interlinking integration

• Main drawbacks
  – Possible loss of information in the conversion
  – Need for an expert to configure the conversion process

Developments still in progress ...