Scuola universitaria professionale
della Svizzera italiana

SUPSI

Interreg
Fondo Europeo di Sviluppo Regionale
UNIONE EUROPEA
ITALIA SVIZZERA - ITALIE SUISSE - ITALIEN SCHWEIZ

GIOCOnDa

From Open Data to Linked Open Data
The GIOCOnDa 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

IARIA

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

(https://www.supsi.ch/isin_en/Research/Educational-technologies-and-open-data.html)

- 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 GIOCOnDa project
- Target groups and data sources
- Methodology to Publish Linked Open Data
- The Conversion Process from Open Data to LOD
- The GIOCOnDa LOD Platform
- Conclusions

Context: the GIOCOnDa 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 **GIOCOnDa 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, NO_2, O_3)

TICINO TURISMO

- Farmhouses
- B&Bs
- Camping
- Hotels
- Hostel
- Alpine huts

OASI - Cantone Ticino

- Environmental data (CO, NO_2, O_3)

Wikidata

- Museums

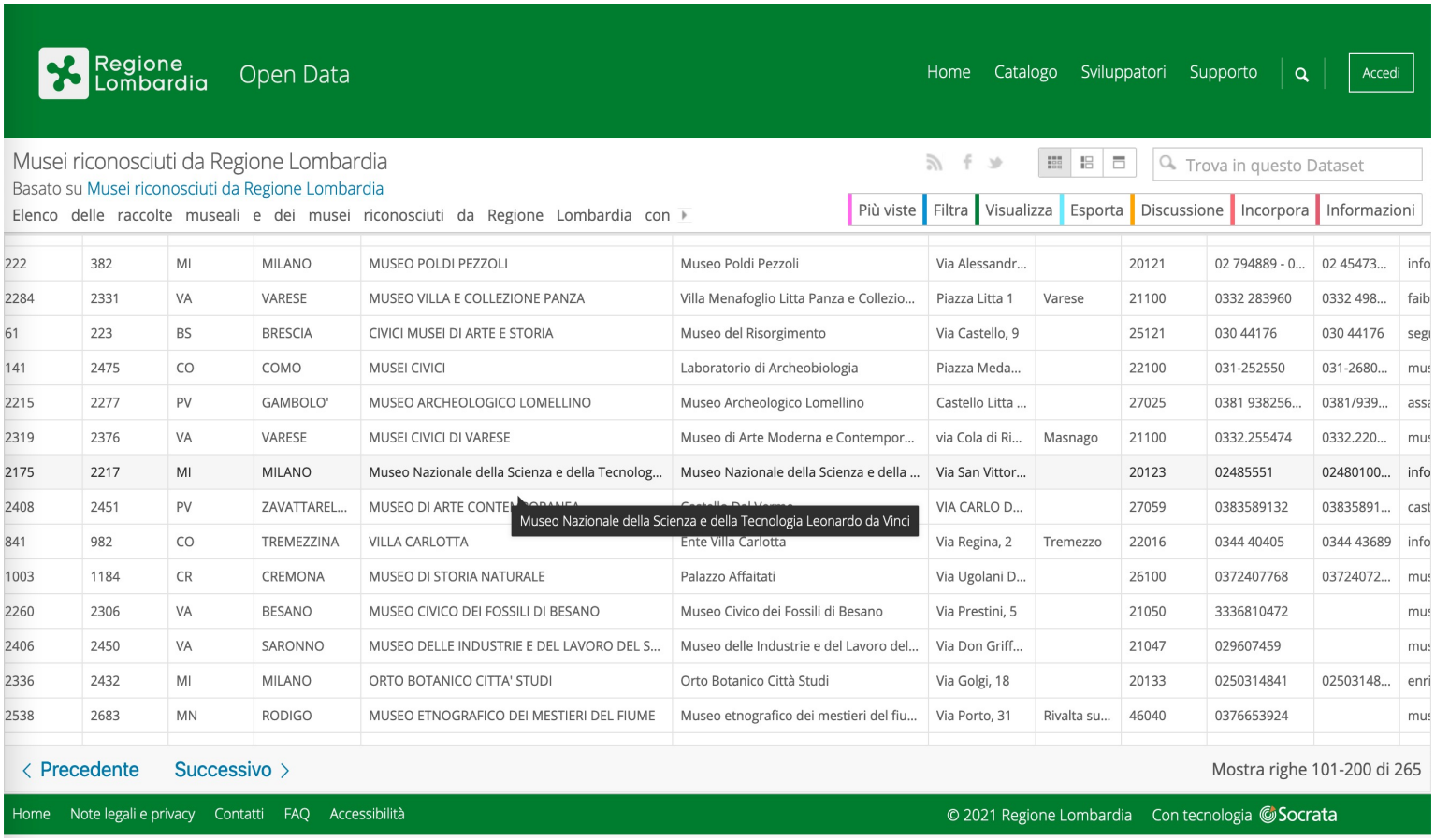
Methodology for LOD publishing

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

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 GIOCOnDa 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

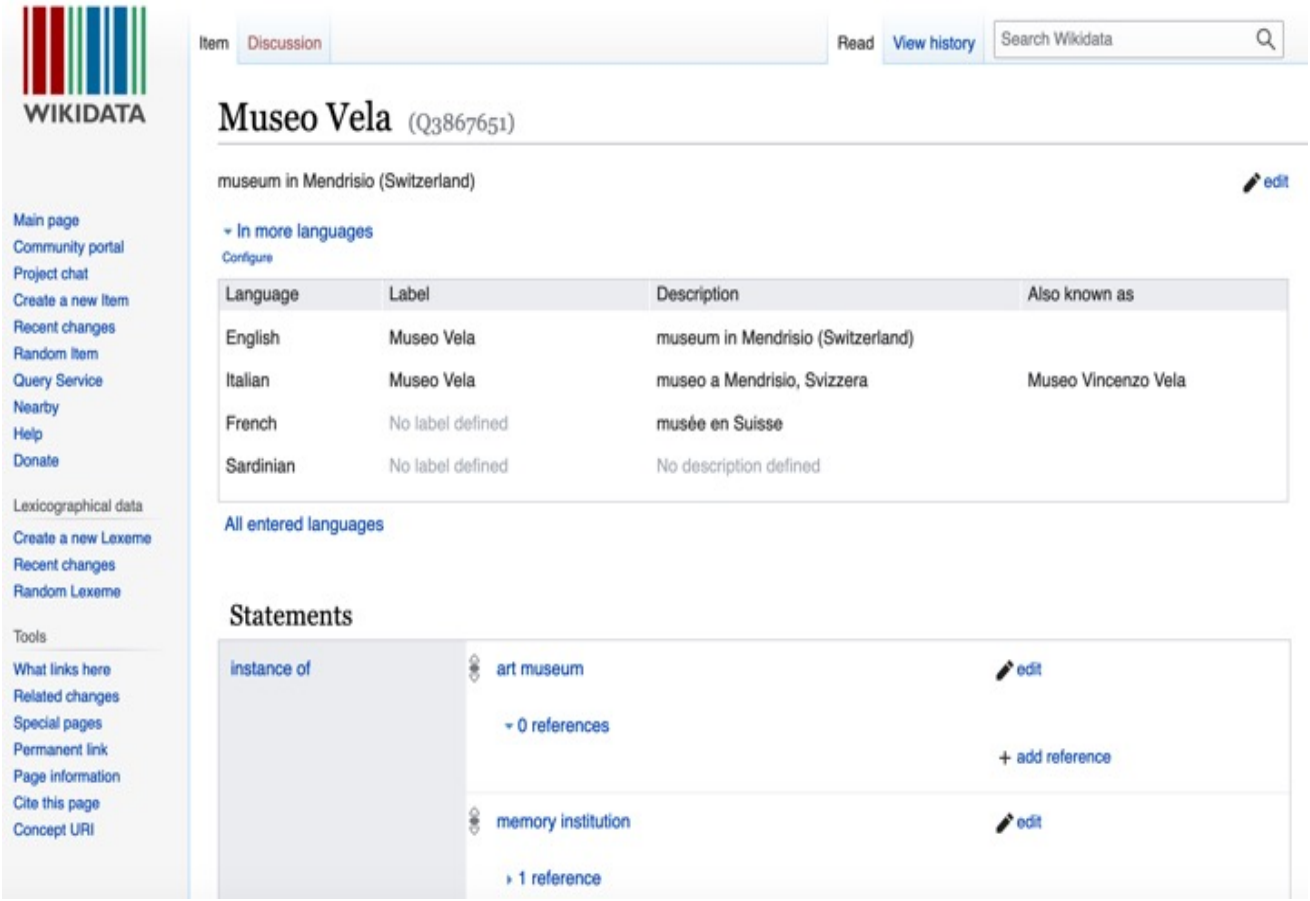
Example: initial format IT museum



222	382	MI	MILANO	MUSEO POLDI PEZZOLI	Museo Poldi Pezzoli	Via Alessandr...		20121	02 794889 - 0...	02 45473...	info
2284	2331	VA	VARESE	MUSEO VILLA E COLLEZIONE PANZA	Villa Menafoglio Litta Panza e Collezio...	Piazza Litta 1	Varese	21100	0332 283960	0332 498...	faib
61	223	BS	BRESCIA	CIVICI MUSEI DI ARTE E STORIA	Museo del Risorgimento	Via Castello, 9		25121	030 44176	030 44176	segr
141	2475	CO	COMO	MUSEI CIVICI	Laboratorio di Archeobiologia	Piazza Meda...		22100	031-252550	031-2680...	mus
2215	2277	PV	GAMBOLO'	MUSEO ARCHEOLOGICO LOMELLINO	Museo Archeologico Lomellino	Castello Litta ...		27025	0381 938256...	0381/939...	assa
2319	2376	VA	VARESE	MUSEI CIVICI DI VARESE	Museo di Arte Moderna e Contempor...	via Cola di Ri...	Masnago	21100	0332.255474	0332.220...	mus
2175	2217	MI	MILANO	Museo Nazionale della Scienza e della Tecnolog...	Museo Nazionale della Scienza e della ...	Via San Vittor...		20123	02485551	02480100...	info
2408	2451	PV	ZAVATTAREL...	MUSEO DI ARTE CONTE...		VIA CARLO D...		27059	0383589132	03835891...	cast
841	982	CO	TREMEZZINA	VILLA CARLOTTA	Ente Villa Carlotta	Via Regina, 2	Tremezzo	22016	0344 40405	0344 43689	info
1003	1184	CR	CREMONA	MUSEO DI STORIA NATURALE	Palazzo Affaitati	Via Ugolani D...		26100	0372407768	03724072...	mus
2260	2306	VA	BESANO	MUSEO CIVICO DEI FOSSILI DI BESANO	Museo Civico dei Fossili di Besano	Via Prestini, 5		21050	3336810472		mus
2406	2450	VA	SARONNO	MUSEO DELLE INDUSTRIE E DEL LAVORO DEL S...	Museo delle Industrie e del Lavoro del...	Via Don Griff...		21047	029607459		mus
2336	2432	MI	MILANO	ORTO BOTANICO CITTA' STUDI	Orto Botanico Città Studi	Via Golgi, 18		20133	0250314841	02503148...	enri
2538	2683	MN	RODIGO	MUSEO ETNOGRAFICO DEI MESTIERI DEL FIUME	Museo etnografico dei mestieri del fiu...	Via Porto, 31	Rivalta su...	46040	0376653924		mus

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

IT museum

```
museum:Museo_2175_sede_2217_Museo_Nazionale_della_Scienza_e_della_Tecnologia_Leonardo_da_Vinci  a
cis:CulturalInstituteOrSite, cis:Museum ;
        rdfs:label "Museo Nazionale della Scienza e della Tecnologia Leonardo da Vinci" ;
        cis:institutionalName "Museo Nazionale della Scienza e della Tecnologia 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 della Scienza e della Tecnologia Leonardo da Vinci";
        cis:siteAddress address:Indirizzo_della_Sede_Museo_scienza_Leonardo_da_Vinci;
        clvapit: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 .
```

CH museum

```
museum:Museo_Q3867651_Museo_Vela    # Q3867651
        a cis:CulturalInstituteOrSite, cis:Museum ;
        rdfs:label "Museo Vela" ;
        cis:institutionalName "Museo Vela" ;
        cis:hasSite site:Sede_Q3867651 ;

        smapit:hasOnlineContactPoint contactPoint:Contatti_Museo_Vela .

site:Sede_Q3867651
        a cis:Site, poiapit:PointOfInterest ;
        rdfs:label "Museo Vela ";
        cis:siteAddress address:Indirizzo_della_Sede_Museo_Vela ;
        clvapit:hasGeometry geometry:geometry_Museo_Vela .
...

contactPoint:Contatti_Museo_Vela
      a smapit:OnlineContactPoint ;
      smapit:hasEmail email:email_museo_Vela;
      smapit:hasTelephoneNumber phone:phone_museo_Vela ;
      smapit:hasWebSite website:web_museo_Vela .
```

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*

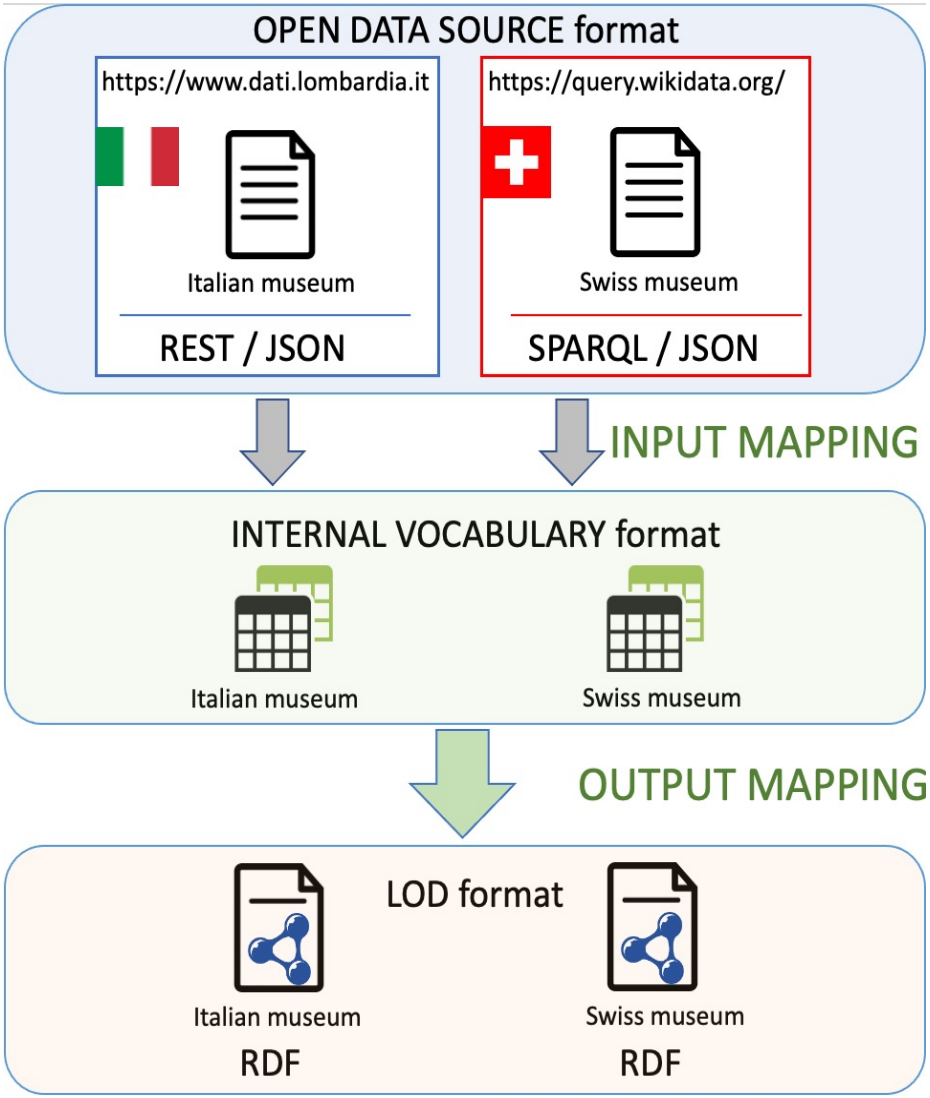
Internal vocabulary

- **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:

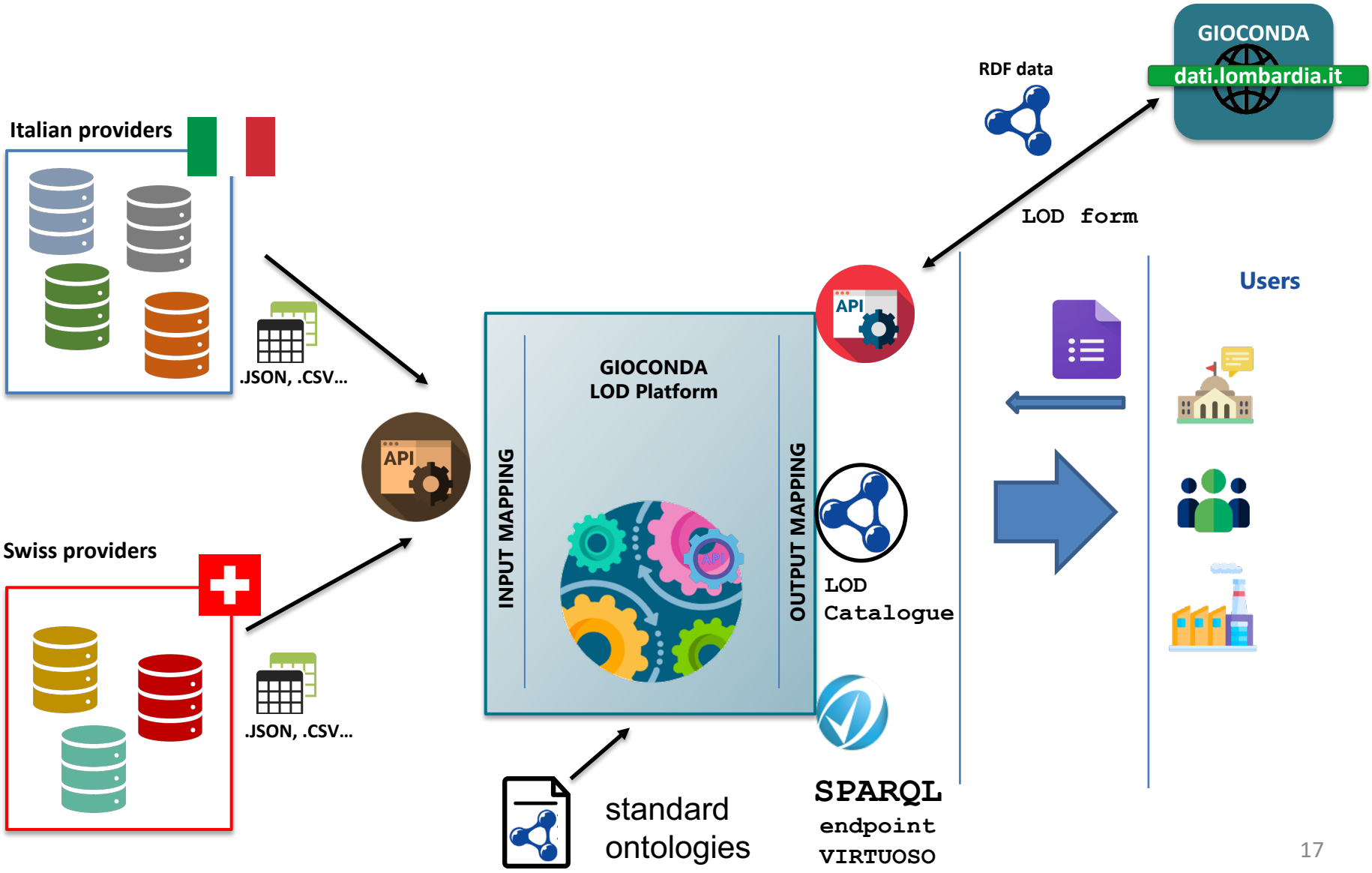
1. **input mapping:** conversion from the input data format to the internal vocabulary
2. **output mapping:** conversion from the internal vocabulary to the ontological LOD format

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

GIOCOnDa LOD Platform

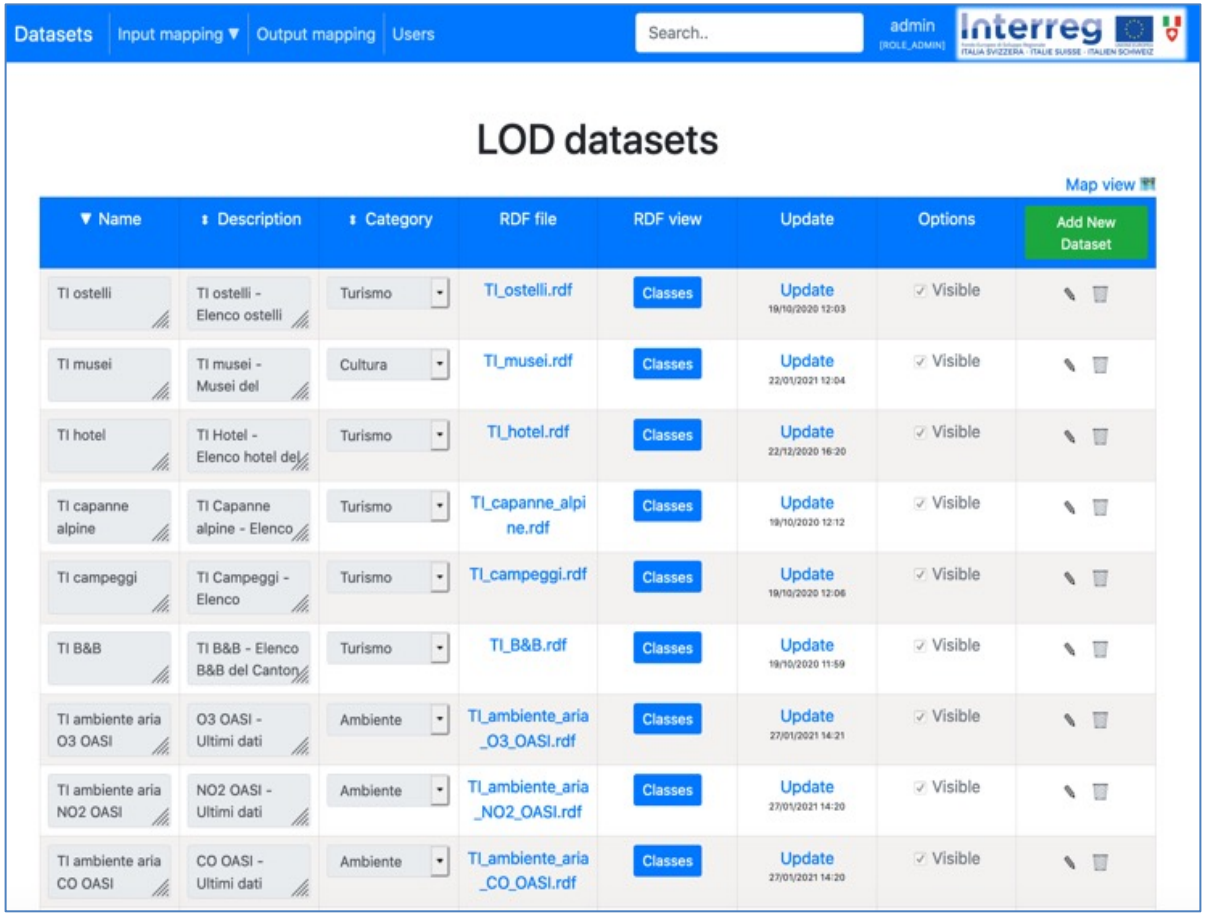
- implemented as a **Java based web application**
- it provides functionalities that enable the publication of LOD datasets and their visualization in a catalogue or in a map

Gioconda LOD architecture

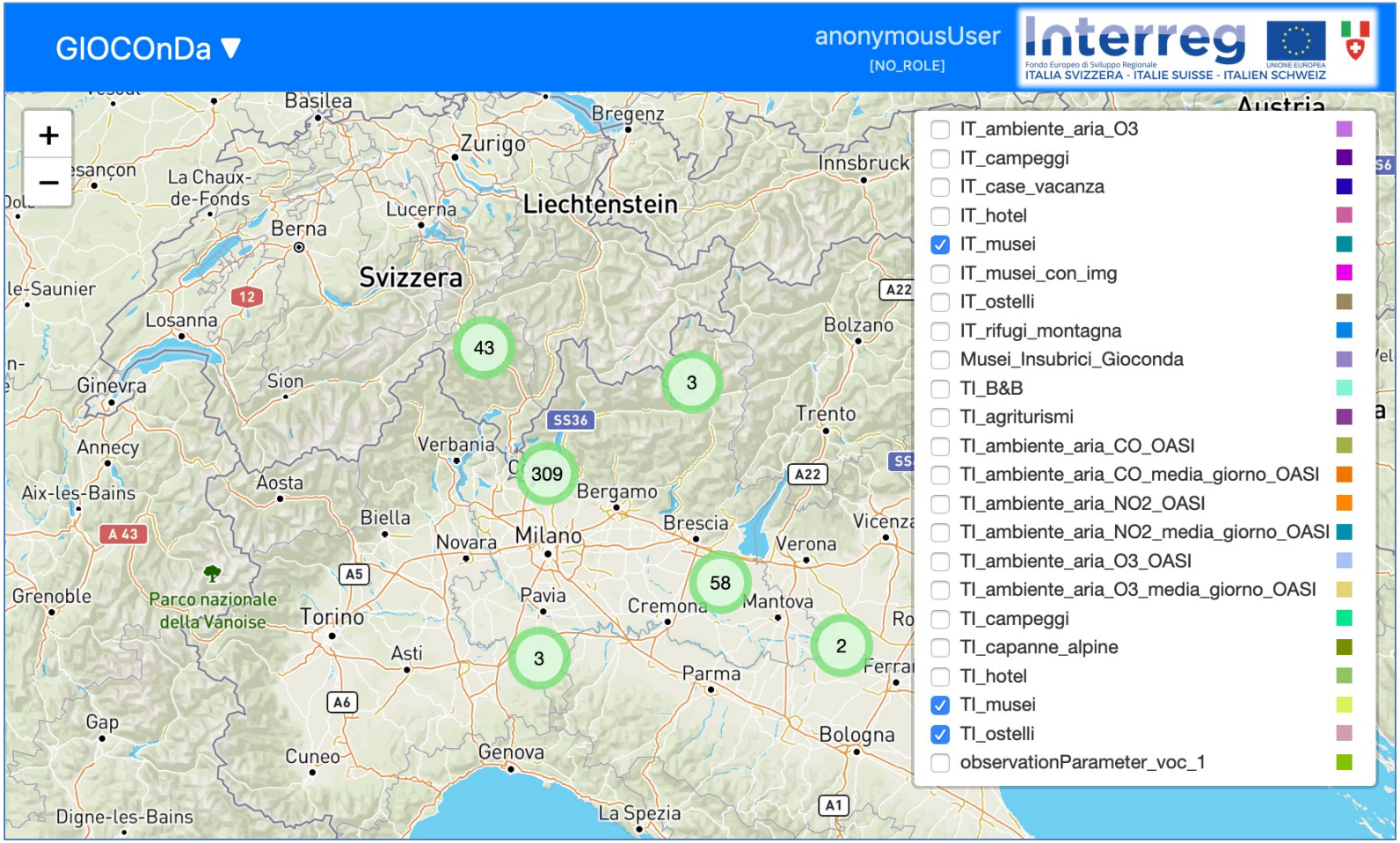


LOD Datasets

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



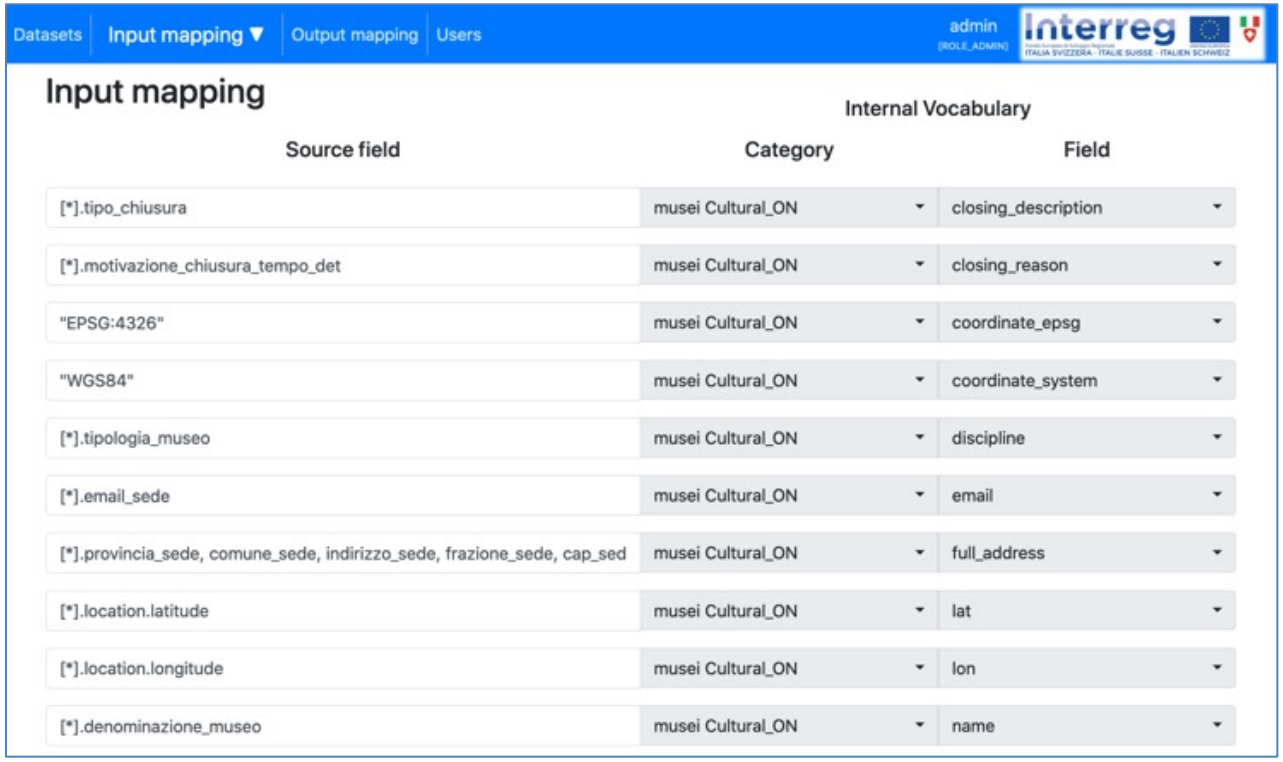
LOD Datasets visualized on the map

Input mapping page

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)

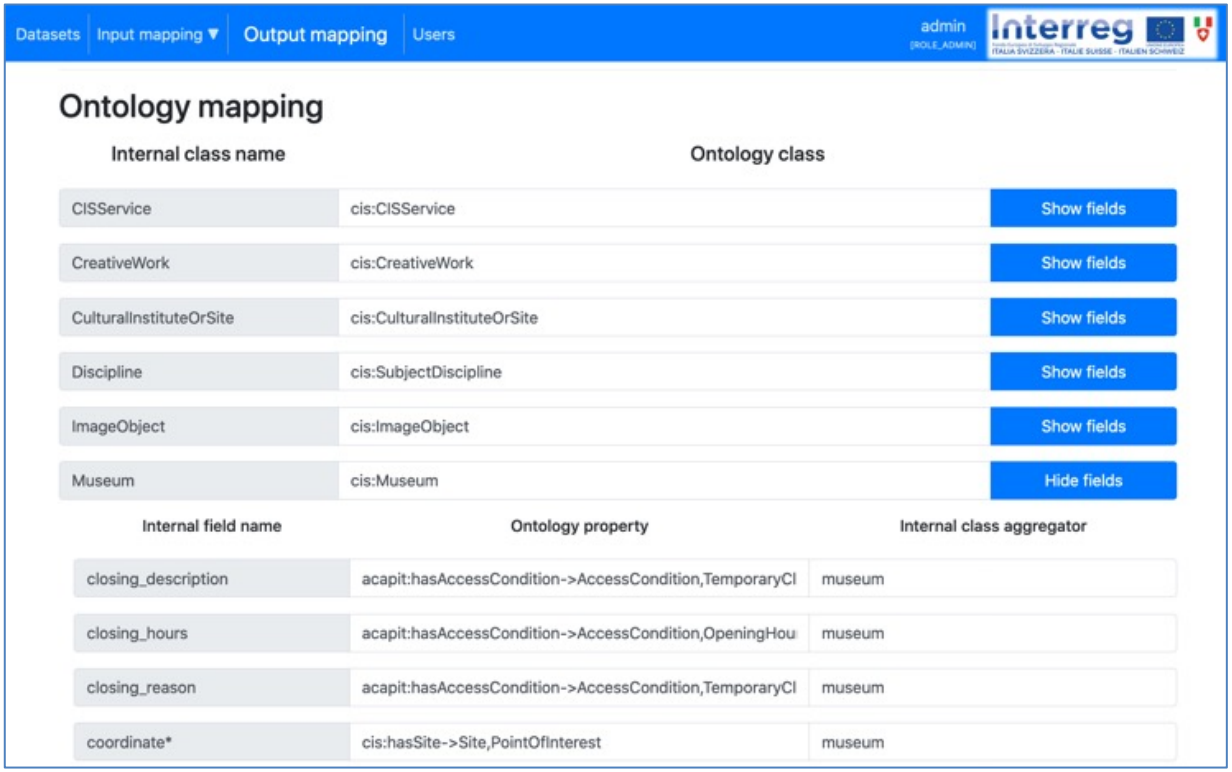


Output mapping page

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 GIOCOnDa LOD platform



Conclusions

- Main advantages of the GIOCOnDa 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 …