SUPSI





### 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

SUPSI

### 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, elearning 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

#### SUPSI

### **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 5



### 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<sub>2</sub>, O<sub>3</sub>)

### TICINO TURISMO

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

#### OASI - Cantone Ticino

• Environmental data (CO, NO<sub>2</sub>, O<sub>3</sub>)

#### 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

2	Regio	one oardia	Open Data			Home Catal	ogo Svilup	opatori S	Supporto Q	Accec	ţi
Muse	riconosc	iuti da Re	gione Lombar	rdia		€ † <i>€</i>		5 Q T	rova in questo آ	Dataset	
Elenco	delle racc	olte musea	ili e dei musei	riconosciuti da Regione Lombardia con	Più viste	Filtra Visuali	zza Esporta	a Discuss	ione Incorpora	Informazi	.oni
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	fail
51	223	BS	BRESCIA	CIVICI MUSEI DI ARTE E STORIA	Museo del Risorgimento	Via Castello, 9		25121	030 44176	030 44176	seg
41	2475	со	сомо	MUSEI CIVICI	Laboratorio di Archeobiologia	Piazza Meda		22100	031-252550	031-2680	mu
2215	2277	PV	GAMBOLO'	MUSEO ARCHEOLOGICO LOMELLINO	Museo Archeologico Lomellino	Castello Litta		27025	0381 938256	0381/939	ass
2319	2376	VA	VARESE	MUSEI CIVICI DI VARESE	Museo di Arte Moderna e Contempor	via Cola di Ri	Masnago	21100	0332.255474	0332.220	mı
175	2217	MI	MILANO	Museo Nazionale della Scienza e della Tecnolog	Museo Nazionale della Scienza e della	Via San Vittor		20123	02485551	02480100	inf
408	2451	PV	ZAVATTAREL	MUSEO DI ARTE CONTEL MUSEO Nazionale della Sci	enza e della Tecnologia Leonardo da Vinci	VIA CARLO D		27059	0383589132	03835891	cas
41	982	со	TREMEZZINA	VILLA CARLOTTA	Ente Villa Carlotta	Via Regina, 2	Tremezzo	22016	0344 40405	0344 43689	inf
003	1184	CR	CREMONA	MUSEO DI STORIA NATURALE	Palazzo Affaitati	Via Ugolani D		26100	0372407768	03724072	mı
260	2306	VA	BESANO	MUSEO CIVICO DEI FOSSILI DI BESANO	Museo Civico dei Fossili di Besano	Via Prestini, 5		21050	3336810472		mı
406	2450	VA	SARONNO	MUSEO DELLE INDUSTRIE E DEL LAVORO DEL S	Museo delle Industrie e del Lavoro del	Via Don Griff		21047	029607459		mı
2336	2432	MI	MILANO	ORTO BOTANICO CITTA' STUDI	Orto Botanico Città Studi	Via Golgi, 18		20133	0250314841	02503148	en
:538	2683	MN	RODIGO	MUSEO ETNOGRAFICO DEI MESTIERI DEL FIUME	Museo etnografico dei mestieri del fiu	Via Porto, 31	Rivalta su	46040	0376653924		mı
< Pre	cedente	Succes	ssivo >	occihilità		@ 2021 Deci		lia Carata	Mostra righe	101-200 di	265

#### Source: <u>https://www.dati.lombardia.it/Cultura/Musei-riconosciuti-da-Regione-</u> Lombardia/3syc-54zf

### Example: initial format CH museum



Source: https://www.wikidata.org/wiki/Q56876232

CH mus

museum:Museo

site:Sede Q38

contactPoin a sma smapi smapi smapi

. . .

The Conversion Process from Open Data to LOD



### **Example:** final RDF format IT and CH museums

	museum:Museo_2175_sede_2217_Museo_Nazionale_de	lla_Scienza_e_della_Tecnologia_Leonardo_da_Vinci a
IT museum	cis:CulturalInstituteOrSite, cis:Museum ;	ienza e della Tecnologia Leonardo da Vinci" -
	cis:institutionalName "Museo Nazional	le della Scienza e della Tecnologia Leonardo da Vinci" ;
	cis:hasSite site:Sede 2217;	,
	_	
	smapit:hasOnlineContactPoint contact	Point:Contatti_Museo_Leonardo_da_Vinci ;
	site-Sede 2217	
	a cis:Site, poiapit:PointOfInterest;	
	rdfs:label "Museo Nazionale della Sc.	ienza e della Tecnologia Leonardo da Vinci";
	cis:siteAddress address:Indirizzo_de	lla_Sede_Museo_scienza_Leonardo_da_Vinci;
	clvapit:hasGeometry geometry:geometry	y_Museo_Leonardo_da_Vinci .
	contactPoint:Contatti Museo Leonardo da Vinci	
	a smapit:OnlineContactPoint ;	
	<pre>smapit:hasEmail email:email_museo_Leonar</pre>	do_da_Vinci;
nuseum	<pre>smapit:hasTelephoneNumber phone:phone_mu</pre>	seo_Leonardo_da_Vinci ;
	smapit:hasTelephoneNumber fax:fax_museo_	Leonardo da Vinci ;
useo_Q3867651_Museo_Vela # Q38676	smapit:naswebsite website:web_museo_heon	ardo_da_vinci .
a cis:CulturalInstituteOrSite, cis:M rdfs.label "Museo Vela" :	useum ;	
cis:institutionalName "Museo Vela" ;		
cis:hasSite site:Sede_Q3867651 ;		Coloction of the <b>Cultural On</b> ontology
and the base of the second second second	A STATE OF STATES AND A STATES	Selection of the <b>Cultural-On</b> ontology
smapit:nasoniinecontactroint con	cactPoint:Contatti_Museo_Vela .	to model the domain
e_Q3867651		
a cis:Site, poiapit:PointOfInterest	2	Search for a <i>match</i> between the
cis:siteAddress address:Indirizzo de	lla Sede Museo Vela :	
clvapit:hasGeometry geometry:geometry	y Museo Vela .	descriptive fields of the II and CH
		museums and the entelogy classes and
		museums and the ontology clusses and
Point:Contatti_Museo_Vela		properties
smapit:OnlineContactPoint;	Vela.	, ,
mapit:hasTelephoneNumber_phone:ph	one museo Vela :	
napit:hasWebSite website;web muse	o Vela .	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:

- 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 ▼	anonymousUser INO_ROLEJ	¢				
	GIOCOnDa LOD					
	Progetto di Cooperazione Italia-Svizzera GIOCOnDa Interreg V-A, ASSE 5 - Rafforzamento della Governance Transfrontaliera 2019-2021 https://interreg-italiasvizzera.eu/progetti/gioconda/					
Benvenuti nel Data a partire	Benvenuti nel portale LOD sviluppato nel progetto Interreg GIOCOnDa per facilitare la pubblicazione di dataset Linked Open Data a partire da Open Data riguardanti la Regione Lombardia e il Canton Ticino.					
I dataset RDF	I dataset RDF esposti sono navigabili in una mappa, presentati in un catalogo e interrogabili mediante un endpoint SPARQL.					
	Finde Durged & Swippe Regionale TraLia SVIZZERA - TTALIE SUISSE - ITALIEN SCHWEIZ					
	Piattaforma GIOCOnDa					



### Gioconda LOD architecture



The GIOCOnDa LOD Platform functionalities

### 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

tasets Input ma	pping 🔻 🛛 Output m	apping Users		Search			
			LOD da	tasets			
▼ Name	Description	Category	RDF file	RDF view	Update	Options	Map view M Add New Dataset
TI ostelli	TI ostelli - Elenco ostelli ///	Turismo •	TI_ostelli.rdf	Classes	Update 19/10/2020 12:03	☑ Visible	ヽ Ⅲ
TI musei	TI musei - Musei del ///	Cultura •	TI_musei.rdf	Classes	Update 22/01/2021 12:04	☑ Visible	1
TI hotel	Ti Hotel - Elenco hotel del	Turismo •	Tl_hotel.rdf	Classes	Update 22/12/2020 16:20	☑ Visible	ヽ Ⅲ
TI capanne alpine ///	TI Capanne alpine - Elenco	Turismo	Tl_capanne_alpi ne.rdf	Classes	Update 19/10/2020 12:12		ヽ Ⅲ
TI campeggi	TI Campeggi - Elenco	Turismo	TI_campeggi.rdf	Classes	Update 19/10/2020 12:06		N II
TI B&B	TI B&B - Elenco B&B del Canton//	Turismo	TI_B&B.rdf	Classes	Update 19/10/2020 11:59	☑ Visible	N II
TI ambiente aria O3 OASI	03 OASI - Ultimi dati	Ambiente •	TI_ambiente_aria _03_0ASI.rdf	Classes	Update 27/01/2021 14:21		N II
TI ambiente aria NO2 OASI	NO2 OASI - Ultimi dati	Ambiente •	TI_ambiente_aria _NO2_OASI.rdf	Classes	Update 27/01/2021 14:20		ヽ Ⅲ
TI ambiente aria CO OASI	CO OASI - Ultimi dati	Ambiente •	TI_ambiente_aria _CO_OASI.rdf	Classes	Update 27/01/2021 14:20	☑ Visible	N II

#### The GIOCOnDa LOD Platform functionalities

### Map view



LOD Datasets visualized on the map

SUPSI

The GIOCOnDa LOD Platform functionalities

### 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)

tasets Input mapping V Output mapping Users		
Input mapping	Internal	Vocabulary
Source field	Category	Field
[*].tipo_chiusura	musei Cultural_ON -	closing_description •
[*].motivazione_chiusura_tempo_det	musei Cultural_ON	closing_reason •
"EPSG:4326"	musei Cultural_ON	coordinate_epsg *
"WGS84"	musei Cultural_ON	coordinate_system •
[*].tipologia_museo	musei Cultural_ON •	discipline -
[*].email_sede	musei Cultural_ON •	email 👻
[*].provincia_sede, comune_sede, indirizzo_sede, frazione_sede, cap_sed	musei Cultural_ON 👻	full_address •
[*].location.latitude	musei Cultural_ON -	lat 👻
[*].location.longitude	musei Cultural_ON •	lon •
[*].denominazione_museo	musei Cultural_ON -	name 👻

SUPSI

The GIOCOnDa LOD Platform functionalities

### 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

atasets Input mapping 🔻 Output map	oping Users		Interreg 💽 😽
Ontology mapping Internal class name	Ontology clas	SS	
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 cla	iss aggregator
closing_description	acapit:hasAccessCondition->AccessCondition,TemporaryCl	museum	
closing_hours	acapit:hasAccessCondition->AccessCondition,OpeningHou	museum	
closing_reason	acapit:hasAccessCondition->AccessCondition,TemporaryCl	museum	
coordinate*	cis:hasSite->Site,PointOfInterest	museum	

SUPSI

### 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 ...