





Experiences in Knowledge Graph Development and Use in Europe

AIST AIRC Artificial Intelligence Seminar

Oscar Corcho

Ontology Engineering Group Centro de I+D en Inteligencia Artificial (Al.nnovation Space) Universidad Politécnica de Madrid



i ocorcho@fi.upm.es

@ocorcho

18/10/2022

Online



Part 1. A presentation from the academic side

WHAT IS A KNOWLEDGE GRAPH?

In an educational/ research context, I would ask the following questions...

- Have you ever heard about the term Knowledge Graph?
- Are you using a KG in your everyday life?
- Which are the most typical KGs available?



What are knowledge graphs?

- When do we start talking about Knowledge Graphs?
 - o Google's announcement of the Google Knowledge Graph in 2012
 - <u>https://www.blog.google/products/search/introducing-knowledge-graph-things-not/</u>
 - "Things, not strings"
 - o Followed by Airbnb, Amazon, eBay, Facebook, IBM, LinkedIn, Microsoft, Uber, etc.



Let's check a video about it



https://youtu.be/mmQI6VGvX-c

- When do we start talking about Knowledge Graphs?
 - Google's announcement of the Google Knowledge Graph in 2012
 - <u>https://www.blog.google/products/search/introducing-knowledge-graph-things-not/</u>
 - "Things, not strings"
 - Followed by Airbnb, Amazon, eBay, Facebook, IBM, LinkedIn, Microsoft, Uber, etc.
- This dates back to earlier efforts done by the knowledge representation & Web communities on the Semantic Web and the Web of Linked Data



- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific american, 284(5), 34-43.
- Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. In Semantic services, interoperability and web applications: emerging concepts (pp. 205-227). IGI Global.
- Plus the emergence of graph databases beyond triple stores



Ingredients for a Knowledge Graph

- Ontologies
 - Explicit representations of the concepts, relations and properties in the domain ("common information model" = "common schema")
 - Often layered and represented in OWL and/or SHACL rules
- Graph-structured data
 - Each entity has a clear identifier (URI, IRI)
 - Often represented in RDF
- Generated from a rich set of sources





Knowledge Graphs. Where are they built from?

Experiences in Knowledge Graph Development and Use in Europe



Part 2. Let's go now into industry...

WHERE ARE KNOWLEDGE GRAPHS BEING USED? AND WHAT FOR?

- Knowledge graphs in the public sector (Ciudades Abiertas)
 - Open data governance (dataset cataloguing, prioritisation)
 - Vocabulary/ontology management
- A knowledge graph for cloud operations (Huawei Research Ireland)
 - $\circ~$ Basis for a chatbot to support site reliability engineers
 - Based on structured data sources (databases) and server logs
- Knowledge graphs and ontologies for improved data governance in a multinational manufacturing company (non-disclosable, yet)
 - Data governance processes based on the DAMA framework
 - Ontology governance processes to generate ontologies for describing reference data
- Plus a few other pointers to ongoing work
 - European Agency for Railways
 - Drugs4COVID

Disclaimers

- These are three use cases (plus ongoing work) where I have been involved as a service provider
 - In ontology governance and engineering
 - $\circ~$ In KG creation and exploitation



 I may have used other cases, but I preferred giving you those where we can discuss insights in more detail

Knowledge and Data Governance

The core message from my presentation

Besides many other cool applications (chatbots, link prediction, question answering systems), knowledge graphs (and the ontologies behind them) are a very useful mechanism for **a proper organisational data governance** nowadays



Case 1. Knowledge graphs in the public sector



Ontology Engineering "on the wild". Open Government

July 2018 – November 2021

INICIATIVA PLATAFORMA DE GOBIERNO ABIERTO, COLABORATIVA E INTEROPERABLE (121/17-SP)



Details at http://www.ciudadesabiertas.es/ (Spanish)

Esta presentación ha sido elaborada en el marco de la iniciativa `Plataforma de Gobierno Abierto, Colaborativa e Interoperable' cofinanciada por el Ministerio de Economía y Empresa, a través de la Entidad Pública Empresarial Red.es, y por los ayuntamientos de A Coruña, Madrid, Santiago de Compostela y Zaragoza y con la cofinanciación del Fondo Europeo de Desarrollo Regional (FEDER), dentro de la `II Convocatoria de Ciudades Inteligentes'.



Fondo Europeo de Desarrollo Regional "Una manera de hacer Europa"

Why do we need ontologies for open data publication?

However, when we provide open data, wouldn't it be good to publish following the same data structures?

Publish	Publish	Publish
Extract	Extract	Extract
lon stop_url		
872 -4.424393 http://ww	link: "http://www.urbanosde:	<pre>zaragoza.es/frm_esquemaparadatime.php?poste=1",</pre>
teEnd Line Label NameA /12/2999 001 1 PLAZA DE CRIS /"ayto:barada";"ayto:sentido";"dc:identifier";"dc:modified"; AAD/2wBDACgcHiMeGSgjISMtKygwFGRBPDc3PHtYXUlkKYC21o+AjIqgtObD	TO REY "gm:coordX";"gm:co boKrarYqMyP/L2u71// title: "(1) Principio de Lín icon: "//Www.zaragoza.es/con - geometry: { type: "Point", - coordinates: [675945,	nea Líneas: 501", ntenidos/iconos/bus.png",
	Image: Non-State State	Image: Stop_url Image: Stop_url Stop_url </td

Develop once, deploy everywhere

Source: Esther Minguela (Localidata)

A bit of history... Open Data Guide by FEMP

DATOS ABIERTOS

Guía estratégica para su puesta en marcha Conjuntos de datos mínimos a publicar



First edition: 2017 Second edition: 2019



http://datos.gob.es/es/noticia/la-femp-elabora-una-guia-de-datos-abiertos-para-ayuntamientos-y-entidades-locales http://femp.femp.es/files/3580-1617-fichero/Guía%20Datos%20Abiertos.pdf https://datos.gob.es/es/documentacion/datos-abiertos-femp-2019-40-conjuntos-de-datos-publicar-por-las-entidades-

locales

http://femp.femp.es/files/3580-1938-fichero/DATOS%20ABIERTOS%20FEMP%202019.pdf

Experiences in Knowledge Graph Development and Use in Europe

A bit of history... FEMP – 40 datasets (2019)



A bit of history... FEMP – 40 datasets (2019)



Up to date information at: https://opencitydata.github.io/CatalogoFEMP/

Red temática española de Open Data y Smart Cities

> **CIUDADES ABIERTAS**

> > **CONCELLO DE SANTIAGO**

UNIÓN EUROPEA

"Una manera de hacer Europa"

Fondo Europeo de Desarrollo Regional

🚼 Zaragoza

https://github.com/opencitydata

CIUDADES CIUDADES E ISLAS INTELIGENTES

red.es

MADRID

Experiences in Knowledge Graph Development and Use in Europe

A bit of history... FEMP data catalogue in GitHub

- Available at
 - o https://github.com/opencitydata
 - <u>https://opencitydata.github.io/Catalo</u> <u>goFEMP/</u> (website)
- Ongoing work on specific vocabularies and dataset transformations (on MSc theses co-supervised by officials from the city of Madrid)





How did we do it? The stakeholders

Honorio Enrique Crespo Díaz-Alejo - Ayuntamiento de Madrid María Carmen Ruiz Moreno - Ayuntamiento de Madrid María Jesús Fernández Ruiz - Avuntamiento de Zaragoza María Jesús Gallego San Miguel - Ayuntamiento de Madrid María del Mar Arribas de Andrés - Ayuntamiento de Madrid Víctor Morlán Plo - Ayuntamiento de Zaragoza Antonio Bermejo Aguña - Ayuntamiento de Madrid Francisco Javier Catalina Celaya - Ayuntamiento de Madrid Andrés Iglesias Pardo – EMT Madrid Andrés Recio Martín – EMT Madrid José Antonio Chanca Cáceres - Ayuntamiento de Zaragoza Ana Bajo Prieto - Ayuntamiento de Madrid Juan Antonio López López - Ayuntamiento de A Coruña Laura Gris Pérez - Ayuntamiento de Madrid Fernando París Roche - Ayuntamiento de Zaragoza Araceli Rollán Arrojo - Ayuntamiento de Madrid Carmen Hervás Bautista - Ayuntamiento de Madrid Manuel Alejandro Pose Pose - Avuntamiento de A Coruña Ángel Pueyo Campos - Universidad de Zaragoza Sergio Valdivielso Pardos - Universidad de Zaragoza Antonio Herrero Martínez – Ayuntamiento de Zaragoza Rubén Notivol Bezares – Ayuntamiento de Zaragoza Francisco Javier Martínez – Ayuntamiento de Zaragoza

. . .

The result. 14 ontologies and 44 thesauri

14 inter-related ontologies (and 44 thesauri) for homogeneous Open City Data publication: Business census, municipal agenda, census datacubes, noise pollution, budget, bicycles, etc.

Vocabulario	Fecha Publicación	Prefijo	Serialización	Licencia	Idioma	Dominio	Enlaces	Descripción
Censo de locales y terrazas, así como sus actividades económicas y licencias de apertura asociadas	16/11/18	escom	rdf+xml html turtle	СС-ВУ	es en	comercio	 repositorio issues requisitos releases webinar 	Vocabulario para la representación de datos sobre el censo de locales y terrazas, así como sus actividades económicas y licencias de apertura asociadas.
Agenda Municipal	21/10/19	esagm	rdf+xml html turtle	СС-ВУ	es en	sector-publico	 □ repositorio ① issues ≃ requisitos ∿ releases ♥ webinar 	Vocabulario para la representación de datos de la agenda municipal que comprende las reuniones de los órganos colegiados y las reuniones en general, actos y reuniones con los medios de comunicación que realiza el alcalde/sa, concejales, directivos y personal eventual con motivo del ejercicio de su cargo.
Padrón de Habitantes s.es/vocabularios/#	04/02/20	espad	rdf+xmi html turtle	СС-ВУ	es en	demografia	 □ repositorio ① issues ~= requisitos ♡ releases 	Vocabulario para la representación de los datos del padrón que provienen de los ficheros de relación de habitantes que intercambian los Ayuntamientos y el Instituto Nacional de Estadística. Estos datos de intercambio corresponden a los microdatos, los datos crudos (raw data) de cada

Some with an associated paper: Ruckhaus E, Antón-Bravo A, Scrocca M, Corcho O (2021-2022) Applying the LOT methodology to a Public Bus Transport Ontology aligned with Transmodel: Challenges and Results. Semantic Web Journal (in press) <u>http://www.semantic-web-</u> journal.net/system/files/swj2841.pdf

Espinoza-Arias P, Poveda-Villalón M;, García-Castro R, Corcho O (2019) Ontological Representation of Smart City Data: From Devices to Cities (2019) Appl. Sci. 9(1):32. https://doi.org/10.3390/app9010032

https://vocab.ciudadesabiertas.es/

Example. Local budget (proposed and actual)



A few recommendations for other lectures...





https://es.slideshare.net/ocorcho/ontology-engineering-atscale-for-open-city-data-sharing https://es.slideshare.net/ocorcho/slowcooked-dataand-apis-in-the-world-of-big-data-the-view-from-acity-perspective

Recommendations for ontology governance

Recommendations for open data governance

Experiences in Knowledge Graph Development and Use in Europe

What about a specific deployment? The city of Zaragoza's knowledge graph

Espinoza-Arias, P.; Fernández-Ruiz, M.J.; Morlán-Plo, V.; Notivol-Bezares, R.; Corcho, O. The Zaragoza's Knowledge Graph: Open Data to Harness the City Knowledge. *Information* **2020**, *11*, 129. <u>https://doi.org/10.3390/info11030129</u>

Where can open data be found in Zaragoza?



www.zaragoza.es

Technical architecture



A knowledge graph: what for?

Conoce y Explora Zaragoza

Development of a data space for the city of Zaragoza

Combined use of thematic maps, dashboards, indicators and data generated by the city council

as well as those created by citizens,

social agents, etc.



Q 🚠

A knowledge graph: what for?



- Open data by default
 - The municipality services are the first reusers of open data
 - The open data portal is not a data graveyard Ο
 - Datasets + APIs \bigcirc
- A proper open data governance
 - Avoid duplication of data across your information systems
- A Knowledge graph inside • Knowledge graph principles are key for this (unique identifiers)
 - Processes and people (multidisciplinar teams are essential)
 - Combine data with maps, dashboards, indicators...
- Homogeneisation
 - Inside the city (common patterns)
 - Across cities (e.g., as in Ciudades Abiertas and on the FEMP catalogue)

Case 2. A knowledge graph for cloud operations (Huawei Research Ireland)

Corcho, O., Chaves-Fraga, D., Toledo, J., Arenas-Guerrero, J., Badenes-Olmedo, C., Wang, M., ... & Zhang, P. (2021, October). A high-level ontology network for ICT infrastructures. In *International Semantic Web Conference* (pp. 446-462). Springer, Cham.



Unknown autor. Available under license CC BY-NC

Motivational example (I)

A sudden drop in performance

(e.g., data throughput) is detected in the provision of a cloud-based product or service (e.g., in a video transmission application that uses a content delivery network).



Unknown autor. Available under license <u>CC BY-NC</u>

Site reliability engineers (SREs) need to be able to inspect the topology of services and microservices (e.g., object storage services, domain name services, elastic cloud servers) that are used by the product where the problem is detected

Motivational example (II)

Services and microservices run in specific servers and clusters and are based on specific software module versions, which are available in some software directory.

Those servers (commonly virtual servers) are running on specific configurations of hardware items (including hardware servers, network cards, etc.) that are hosted in a data center.



Unknown author. Available under license CC BY-NC

Unknown author. Available under license CC BY

All this should be done independently of which specific types of infrastructures, hardware and software providers are being used.

We need to build an ontology (or better an ontology network) for this domain.

Information sources for this:

- Configuration Management Database models (e.g., BMC CMDB)
- IT Service Management Systems
- IT Asset Management databases and tools

Including database models available and in use by Huawei

- 85 tables for entities, 59 tables for views
- Some sample data for the main entities



Ontology network / catalogue



Experiences in Knowledge Graph Development and Use in Europe

Case 3. Knowledge graphs and ontology engineering for improved data governance in a manufacturing Company



Knowledge graphs and ontology engineering in a manufacturing company

- Two teams/departments: Data Governance and Ontology Governance
 - Difficulties in handshaking when ontologies are being delivered to the data management team





Governance creates the coordination framework for these components to work together



Experiences in Knowledge Graph Development and Use in Europe

Ontology Governance Model and workstreams



Governance Principles





Experiences in Knowledge Graph Development and Use in Europe

Governance Principles



Recommendations

- Rec 1 on IRIs
- Rec 2 on Metadata
- Rec 3 on Format
- Rec 4 on Reuse
- Rec 5 on Naming Conventions
- Rec 6 on Licensing



Experiences in Knowledge Graph Development and Use in Europe

Principles and the ontology and knowledge graph lifecycle



Standards Journey



- Knowledge graphs in the public sector (Ciudades Abiertas)
 - Open data governance (dataset cataloguing, prioritisation)
 - Vocabulary/ontology management
- A knowledge graph for cloud operations (Huawei Research Ireland)
 - $\circ~$ Basis for a chatbot to support site reliability engineers
 - Based on structured data sources (databases) and server logs
- Knowledge graphs and ontologies for improved data governance in a multinational manufacturing company (non-disclosable, yet)
 - Data governance processes based on the DAMA framework
 - Ontology governance processes to generate ontologies for describing reference data
- Plus a few other pointers to ongoing work
 - European Agency for Railways
 - o Drugs4COVID

- Plus a few other pointers to ongoing work
 - European Agency for Railways: homogeneise data about rail infrastructure in Europe, run compatibility checks between vehicles and tracks, generate routebooks, etc.
 - <u>https://data-interop.era.europa.eu/</u>
 - <u>https://data-interop.era.europa.eu/era-vocabulary/</u>

EUROPEAN MACHINE FOR MALIWAYS					₀ Route Compatibility Check	Q Search	😵 Status	븆 Endpoint	A7 ¥
Route C	Compatibility Ch	ieck	i	+ ea - is batteros	Huesca		Tremp		3
From: 📁 B/	ARCELONA-CAN	UNIS - ES	71901 × ~	· •	Barbastro				t i
+ Via point	ıt				Monzón	4	115	J. Jos	X
To: 🗖 CIM	DE ZARAGOZA - I	S70803	× ~			1ª			
11-001-000	02-1-001 - Euro 44)00 Type II	\times \vee			Lleida	-		
						3/1			
Route 1 (39	96.348 km):		^		and and	T.C.			
1 BARC	CELONA-CAN TUNI	S (freight to	erminal)		So-von marken	3		Valls	1
Track: Vehicl	:: : e Type: 11-001-0002 [.]	1-001			A CONTRACT		معر	mo-	
Pr	roperties Compati	ole Track	Vehicle	Stan Sand	Alcañiz	12			
Link	Type of train detection system	n no data n no data	G1 • track circuit • wheel detector • loop			Tortosa	2		
· - · ·	Existence of			(©)mepbex		Amposta		52	© Mapbox ©

S

- o Drugs4COVID
 - https://drugs4covid.oeg.fi.upm.es/
 - You can play with our QA system:
 - https://drugs4covid.oeg.fi.upm.es/services/qa

DRUGS4COVID	SERVICES	KNOWLEDGE GRAPH					۹
What traditional medicine has been used to treat Covid-19?					Ø]	
Max Answers					3		
	🗹 D4c 🗌 DB Pedia 🗌	Wikidata					
What traditional medicine has been used to tr herbal formulae	eat Covid-19?						
Type: literal							
Confidence: 0.4							
In the process of treating for COVID-19, the treatments are individu Syndrome differentiation individually is used to guide the selection drugs and herbal formulae also were recommended for treating O2 therapeutic effect of traditional Chinese medicine warrants further	ualized according to traditional n of traditional Chinese medicin OVID-19 in the latest version of r investigation in future studies	Chinese medicine syndrome te treatment using traditional China guideline for diagnosis	diagnosis and Chinese medic s and treatment	treatment (Bian Zhi ine <mark>herbal formulae</mark> : of COVID-19. 3 Ho	eng Lun Zhi). . 62 Some wever, the		



Take home message



*Take home message Esta foto de Autor desconocido está bajo licencia CC BY-NC-ND

- Heterogeneous data sources in data lakes/data governance need homogeneisation
 - $\circ~$ Ontologies and vocabularies provide for that
 - They support reference data representation
- Knowledge graphs provide an alternative for improved exploitation
- This is a necessary first step towards doing more advanced exploitations









Experiences in Knowledge Graph Development and Use in Europe

AIST AIRC Artificial Intelligence Seminar

Oscar Corcho

Ontology Engineering Group Centro de I+D en Inteligencia Artificial (Al.nnovation Space) Universidad Politécnica de Madrid



i ocorcho@fi.upm.es

@ocorcho

18/10/2022

Online