

Provincepedia

The encyclopedia of Spanish provinces

Cesare Curcio
Filippo Calzavara



UNIVERSIDAD
POLITÉCNICA
DE MADRID

Dataset

- 3 datasets initially chosen:
 - Public debt for every Spanish city
 - Population series records through last 10 years
 - Crimes reported by each province
- From these 3 dataset, other 3 refined datasets where derived using OpenRefine:
 - Autonomous community
 - Provinces
 - Type of felonies committed



Provincepedia
Dataset modification

Original Public Debt Dataset

 GOBIERNO DE ESPAÑA MINISTERIO DE HACIENDA				SECRETARÍA DE ESTADO DE HACIENDA			
				SECRETARÍA GENERAL DE FINANCIACIÓN AUTONÓMICA Y LOCAL			
Ejercicio	Código CC.AA.	CC.AA.	Código Provincia	Provincia	Código Municipio	Municipio	Deuda viva 31/12/2017 (miles de euros)
Ejercicio	cdcdad	nomcdad010	cdprov	nomprov010	cdcorp	nomcorp	ImpDeudViva
2017	01	ANDALUCIA	04	ALMERIA	001	Abla	310
2017	01	ANDALUCIA	04	ALMERIA	002	Abrucena	16
2017	01	ANDALUCIA	04	ALMERIA	003	Adra	15,452
2017	01	ANDALUCIA	04	ALMERIA	004	Albánchez	61
2017	01	ANDALUCIA	04	ALMERIA	005	Alboloduy	0
2017	01	ANDALUCIA	04	ALMERIA	006	Albox	20,390
2017	01	ANDALUCIA	04	ALMERIA	007	Alcolea	235
2017	01	ANDALUCIA	04	ALMERIA	008	Alcóntar	149
2017	01	ANDALUCIA	04	ALMERIA	009	Alcudia de Monteagud	24
2017	01	ANDALUCIA	04	ALMERIA	010	Alhabia	2,601
2017	01	ANDALUCIA	04	ALMERIA	011	Alhama de Almería	2,130
2017	01	ANDALUCIA	04	ALMERIA	012	Alicún	189

Implemented public debt dataset

- Only the most recent data available (2017) was used
- The total amount per province has not been used
- A new field called "capital" was added to mark the province capital
- All the other fields have been removed

prov_code	capital	town_name	debt
1	0	Alegr�a-Dulantzi	639
1	0	Amurrio	244
1	0	Aramaio	0
1	0	Artziniega	0
1	0	Armi��n	0
1	0	Arratzua-Ubarrundia	0
1	0	Asparrena	387
1	0	Ayala/Aiara	0
1	0	Ba��os de Ebro/Ma��tueta	64
1	0	Barrundia	0
1	0	Berantevilla	0
1	0	Bernedo	51
1	0	Campezo/Kanpezu	0
1	0	Zigoitia	0
1	0	Kripan	216
1	0	Kuartango	0
1	0	Elburgo/Burgelu	0
1	0	Elciego	29
1	0	Elvillar/Bilar	142
1	0	Iruraiz-Gauna	0
1	0	Labastida/Bastida	694
1	0	Logro��o	0

Original population dataset

[illegible]

Implemented population dataset

year	prov_code		men	women	
2018		2	195009	195357	
2017		2	195681	195893	
2016		2	196394	196564	
2015		2	197389	197539	
2014		2	198256	198428	
2013		2	199828	199682	
2012		2	201085	200505	
2011		2	200830	200016	
2010		2	200366	199081	
2009		2	199799	198082	
2008		2	198260	196261	
2018		3	915455	930089	
2017		3	914824	927748	
2016		3	915629	926816	
2015		3	916055	926119	

- Only data regarding the last 10 years were considered
- National population have been eliminated because not useful
- A new field was added to identify the province of the town
- The total sum of people has been removed because redundant

Original crime dataset

TERRITORIO;TIPOLOGÍA PENAL;2015;2016; Var. % 16/15	
ANDALUCÍA;1.-DELITOS Y FALTAS (EU);258032;252684;-2	1
ANDALUCÍA;2.-HOMICIDIOS DOLOSOS Y ASESINATOS CONSUMADOS (EU);48;49;2	1
ANDALUCÍA;3.-ROBO CON VIOLENCIA E INTIMIDACIÓN (EU);5423;5686;4	8
ANDALUCÍA;4.-ROBOS CON FUERZA EN DOMICILIOS (EU);13571;13945;2	8
ANDALUCÍA;5.-SUSTRACCIÓN VEHÍCULOS A MOTOR (EU);6669;6600;-1	0
ANDALUCÍA;6.-TRÁFICO DE DROGAS (EU);2469;2444;-1	0
ANDALUCÍA;7.-DAÑOS;25514;25882;1	4
ANDALUCÍA;8.-HURTOS;84504;84445;-0	1
Provincia de Almería;1.-DELITOS Y FALTAS (EU);23562;23277;-1	2
Provincia de Almería;2.-HOMICIDIOS DOLOSOS Y ASESINATOS CONSUMADOS (EU);4;5;25	0
Provincia de Almería;3.-ROBO CON VIOLENCIA E INTIMIDACIÓN (EU);511;492;-3	7
Provincia de Almería;4.-ROBOS CON FUERZA EN DOMICILIOS (EU);1974;2067;4	7

Implemented crime dataset

prov_code		type	2015	2016
1	1	1	9414	8947
1	2	2	1	2
1	3	3	144	148
1	4	4	380	274
1	5	5	68	77
1	6	6	26	30
1	7	7	922	944
1	8	8	3506	3568
2	1	1	9398	8779
2	2	2	2	1
2	3	3	158	167
2	4	4	561	516
2	5	5	138	140

- Only the most recent data collected have been considered
- The variation from 2015 to 2016 have been removed
- Only the total crimes reported in the province was considered
- The territory field was transformed in the code of the province in which the crime happened

Additional changes

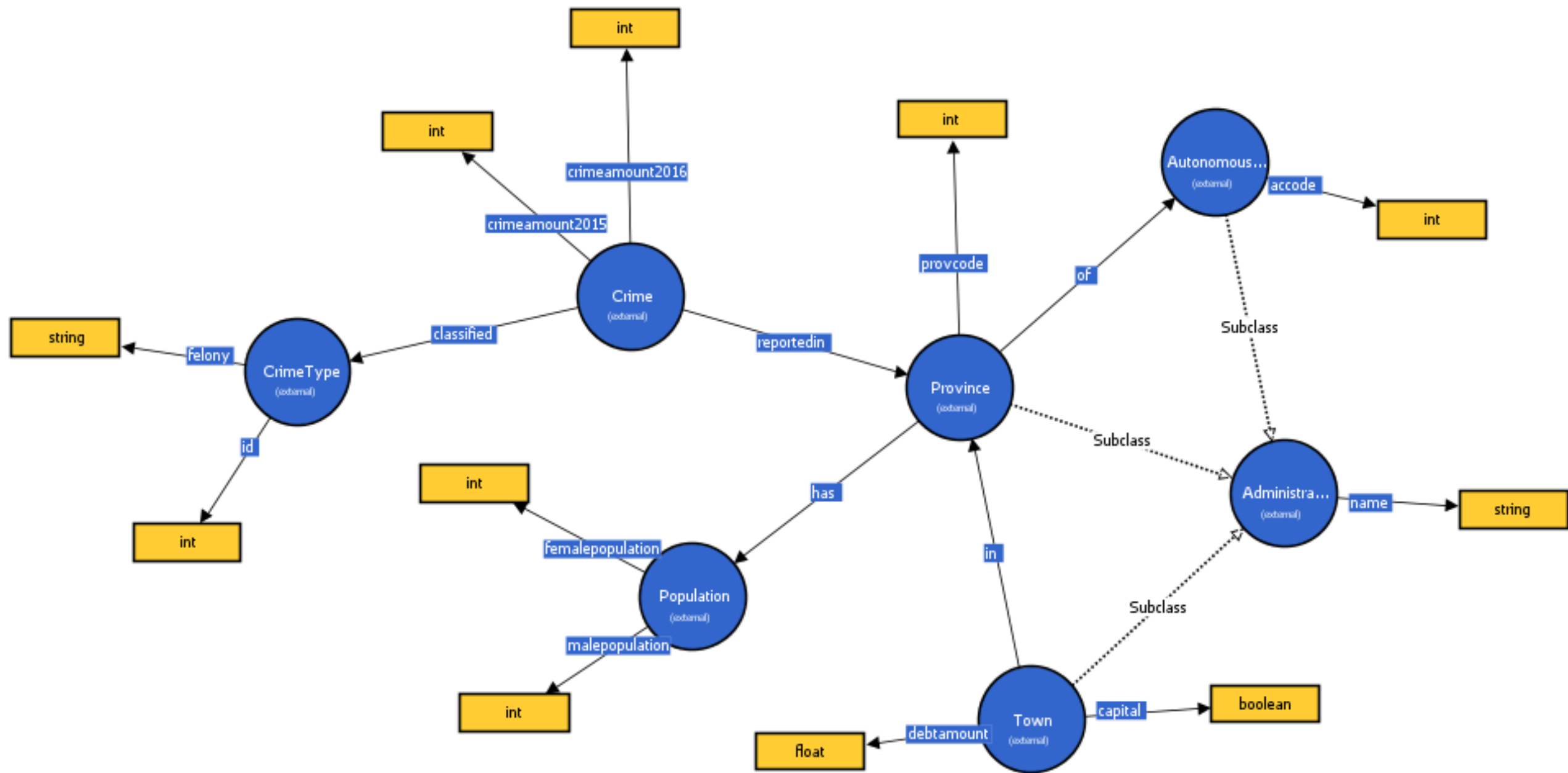
- Void cells was eliminated
- Malformed data was recovered and cleaned or if that was not possible removed
- To link the data of the different datasets, 3 more have been created:
 - The "ac" dataset links the name of each autonomous community to its code
 - The "felonies" one connects the codes of the crime to their name and description
 - The "prov" dataset is used to correlate the name and the id of each province to its corresponding autonomous community

Provincepedia
Ontology creation

Ontology structure

- Implemented in Protégé:
 - 7 classes
 - 5 object properties
 - 11 data properties





Provincepedia

A web application

Conclusions

- A web application based on node.js was developed
 - Connected to WikiData and DBPedia. Used library wikidata-sdk, dbpedia-sparql-client and rdfstore
- GitHub was used for code versioning. A descriptive readme with installation instruction can be read on the repository
- Source code is present in the app folder
- The project changed during the work:
 - Despite the initial idea of an encyclopaedia of all the cities, the datasets found were more suited for an encyclopaedia of provinces
 - Thus the project was changed accordingly

Provincepedia

Thank you!



<https://github.com/philosss/Provincepedia>