

# Grafos de Conocimiento para relaciones clientelares: una propuesta de proyecto colectivo

Objetivo

Tecnología

Ejemplo de proceso

Otros proyectos

No todo el monte es orégano

Posibles tareas y grupos de trabajo

Ejercicios / Preguntas / Debate / Trabajo en grupo

# Mikel Egaña Aranguren

<https://mikel-egana-aranguren.github.io/>

Profesor en UPV/EHU, Grado Informática - Ingenieros Bilbo

Experiencia trabajando con Grafos de Conocimiento (Knowledge Graphs):

- Contratado en empresas  
- Proyectos de investigación académicos
- Proyectos de consultoría para entidades públicas y privadas

# Objetivo

Integrar y publicar datos sobre posibles  
relaciones clientelares en Euskadi de manera  
colaborativa y tecnológicamente sostenible

# Posibles relaciones clientelares

Relaciones que incluyen ...

- Agentes (Políticos, empresarios, funcionarios, ...)
- Entidades (Gobierno, sociedades, servicios, partidos políticos, empresas, ...)

... que podrían representar un conflicto de interés a investigar/documentar/publicar

# Euskadi (CAV)

Cercanía

Dominio bien delimitado administrativamente (Autonomía, Diputación, ...)

Al “Oasis vasco” se le ven las grietas cada vez más

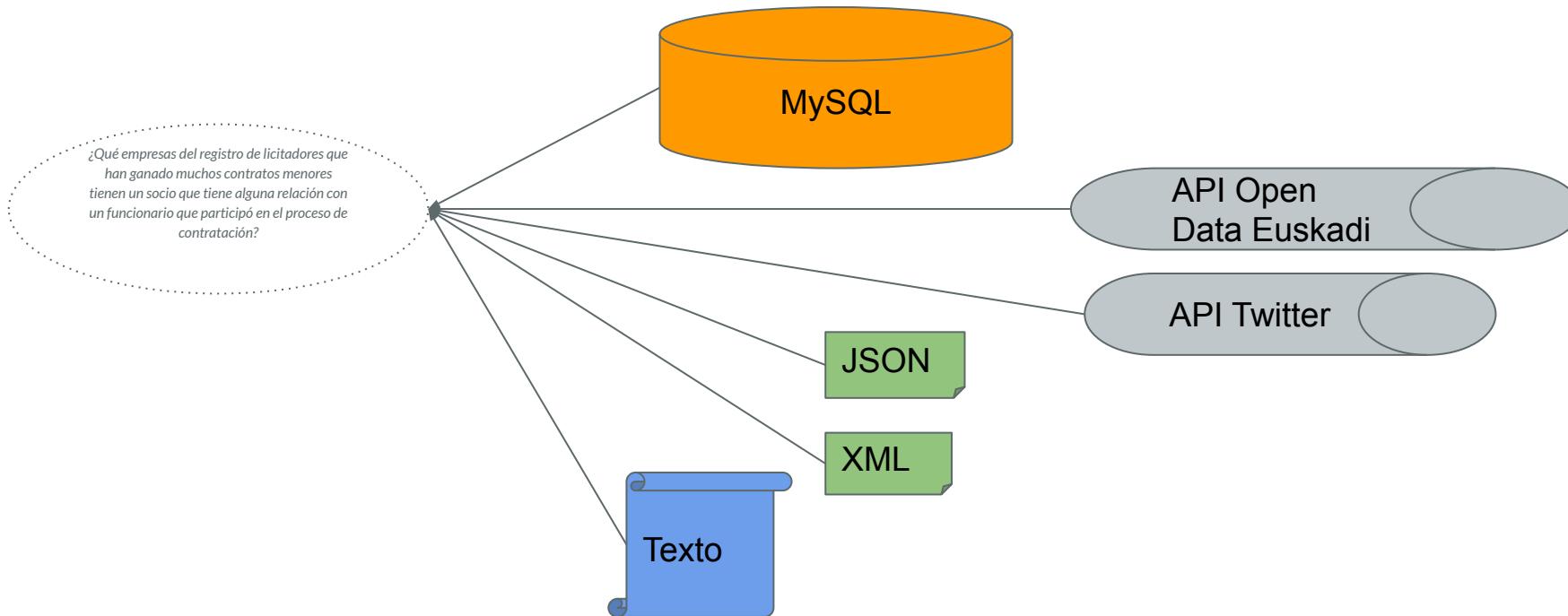
Gran tradición de crítica a las élites y análisis de corrupción: Ahoztar Zelaieta, Hordago, Bilbo Data Lab, ...

# Integrar

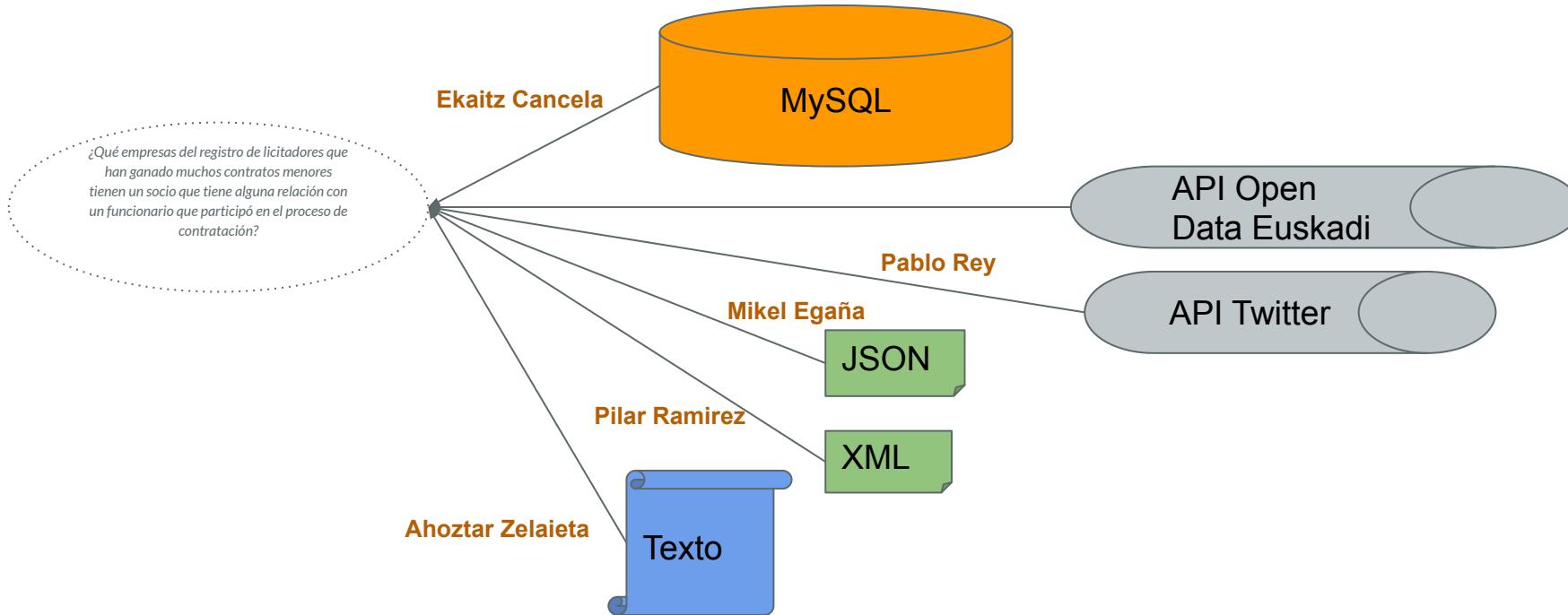
Los datos aislados valen más cuando los integramos con más datos para obtener conclusiones interesantes (Deuda de integración ~ deuda técnica)

*¿Qué empresas del registro de licitadores que han ganado muchos contratos menores tienen un socio que tiene alguna relación con un funcionario que participó en el proceso de contratación?*

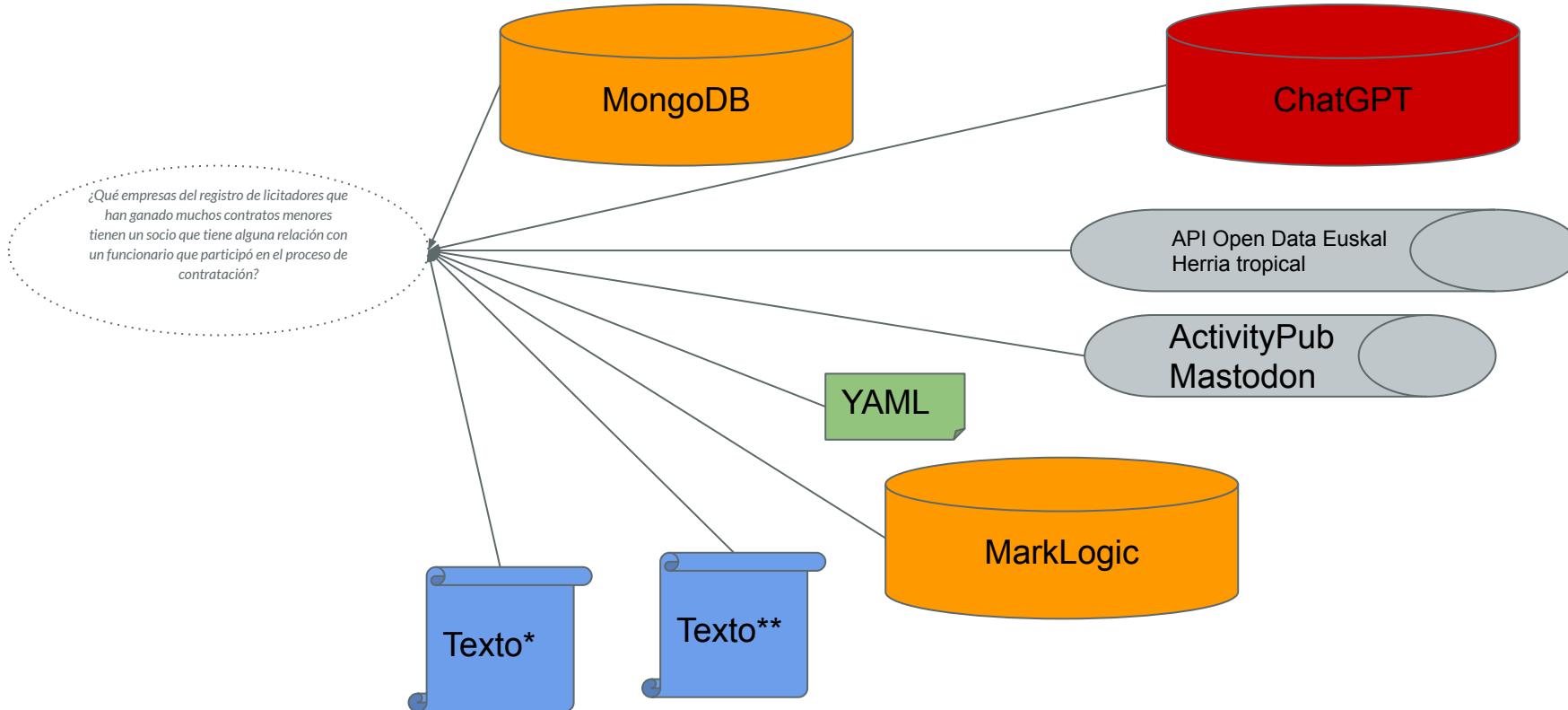
# Integrar

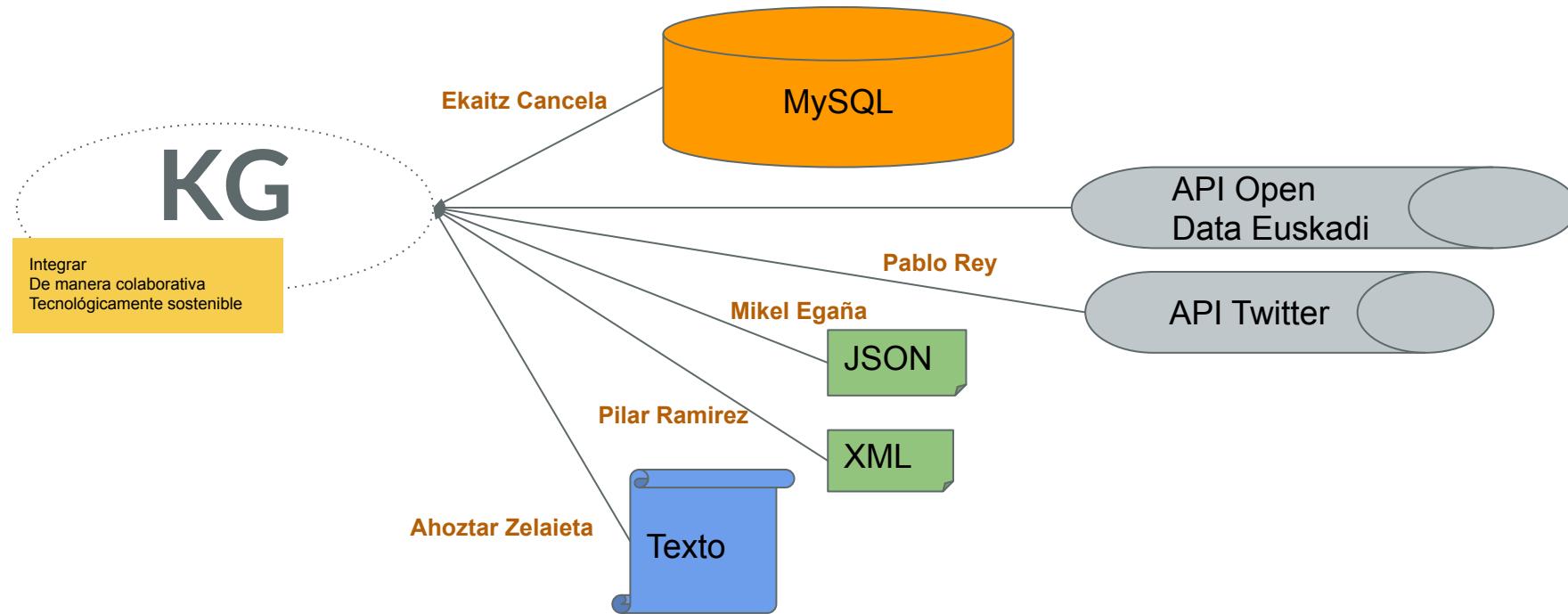


# De manera colaborativa

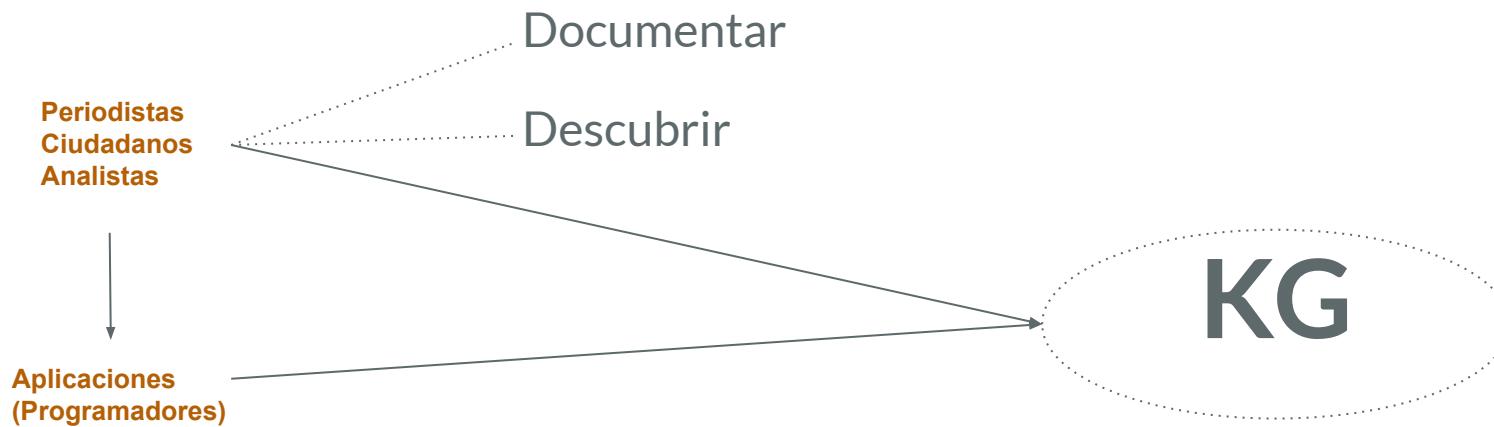


# Tecnológicamente sostenible





# Publicar



mikel-egana-aranguren / BasqueCountryInstitutionsTransparentRelationsGraph (Public)

Code Issues 14 Pull requests Discussions Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file > Code

mikel-egana-aranguren No profile

People No profile 5 days ago

RegistroLicitadoresYEmpresasClasif... Improve README and add People example 3 weeks ago

.gitignore First conversion to RDF 3 weeks ago

LICENSE Initial commit last month

README.md Update README with disclaimer last week

blazegraph.jar Move blazegraph jar upwards 3 weeks ago

README.md

## Basque Country Institutions Transparent Relations Graph

RDF Knowledge Graph for Transparent relations involving Basque Country institutions. The aim of this project is to build a graph to integrate information about entities and individuals that might have a conflict of interest, in order to analyse such information.

DISCLAIMER: This graph integrates only publicly available information from legitimate sources (Open Data portals, news, etc). Therefore the inclusion of a person or entity in this graph is not intended to suggest or imply that they have engaged in illegal or improper conduct.

Data sources

About

Transparent relations involving Basque Country institutions

Readme Apache-2.0 license 0 stars 1 watching 0 forks

Releases

No releases published Create a new release

Packages

No packages published Publish your first package

Languages

Python 100.0%



**DISCLAIMER:** This graph integrates only publicly available information from legitimate sources (Open Data portals, news, etc). Therefore the inclusion of a person or entity in this graph is not intended to suggest or imply that they have engaged in illegal or improper conduct.

KG

Posibles conflictos de interés → “Legales pero ilegítimos” → publicar/analizar

Información pública reusable y publicable

Información “No publicable” → Ilegales → Investigar?Denunciar?

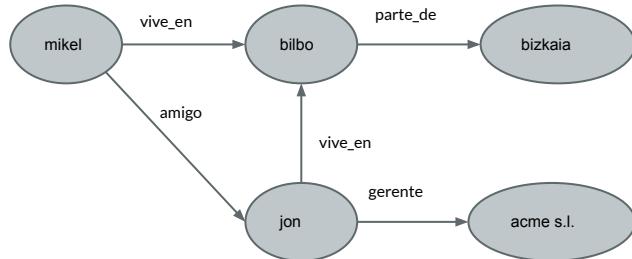
# Tecnología

# Knowledge Graphs

Graph: Modelo de datos basado en nodos y vértices

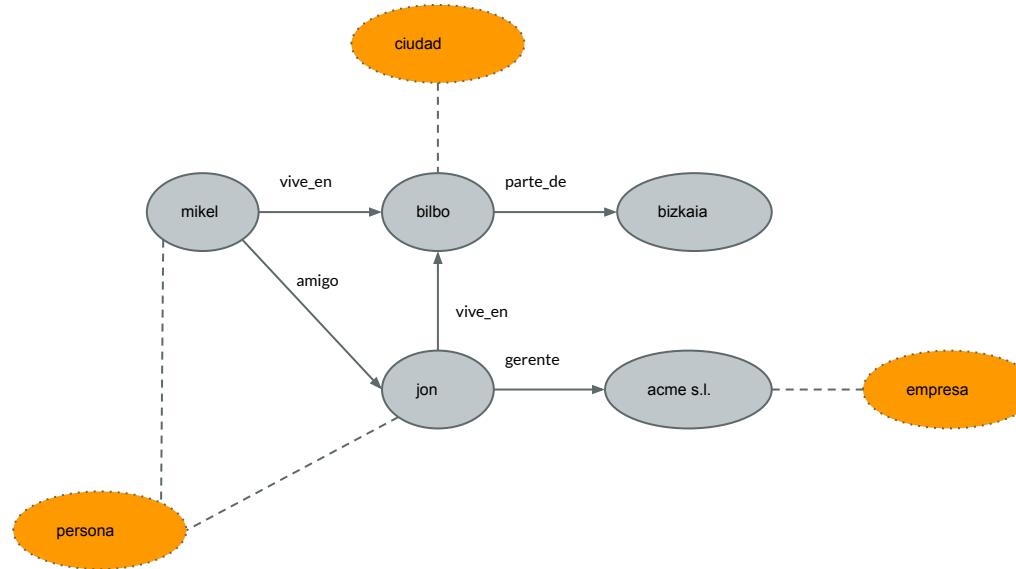
Tripletas formada por dos nodos y un vértice

Muchas tripletas: grafo



# Knowledge Graphs

Knowledge: vocabulario común para definir las entidades del grafo



# Knowledge Graphs

- (+) Modelo extremadamente simple: enlazar grafos e integrar datos con facilidad
- (+) Muy intuitivo
- (+) Centrado en relaciones
- (-) Todo se puede modelar: complejidad en consultas, mantenimiento, etc.
- (-) Ideas muy antiguas en tecnología relativamente nueva (Aunque ya en producción): curva de aprendizaje

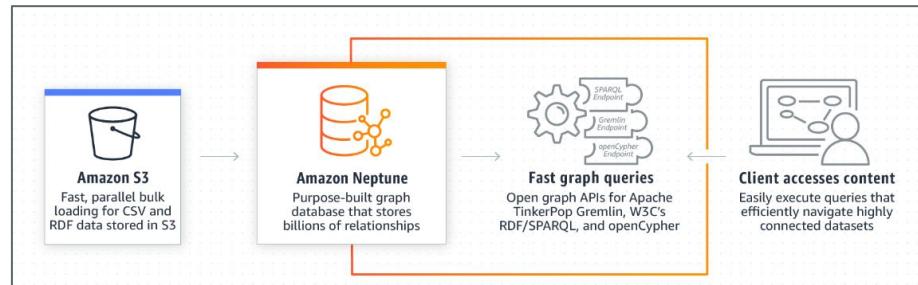
# Knowledge Graphs

The screenshot shows the Schema.org JSON-LD documentation for the Person schema type. It includes a table of properties with their expected types and descriptions, such as address, name, and birthDate.



The screenshot shows a Yandex search result for "Henry Rollins". On the right side of the search results, there is a "knowledge panel" for Henry Rollins, which includes his photo, a bio, and links to his music and tour information.

The screenshot shows the Google Cloud Enterprise Knowledge Graph overview page. It highlights features like Entity Reconciliation API, which merges multiple data sources, and provides a video tutorial titled "How to use Enterprise Knowledge Graph to bridge data silos".



# Knowledge Graphs



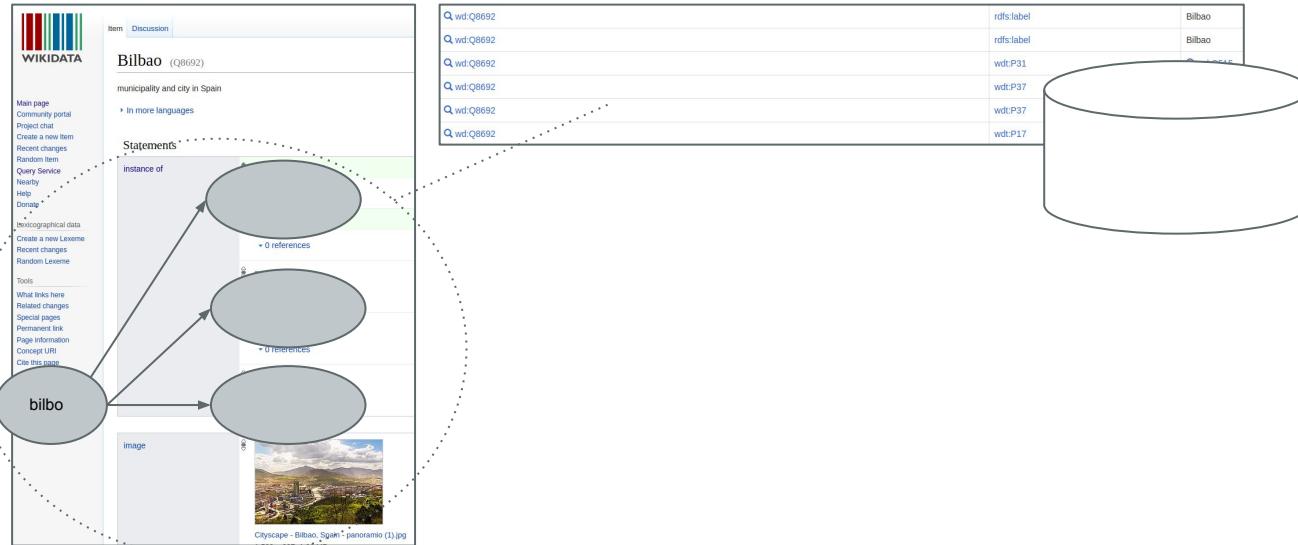
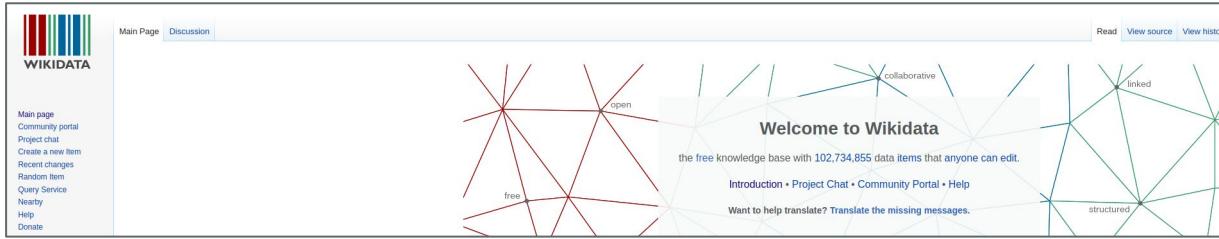
Katariina Kari  
Aug 24, 2022 · 5 min read · Listen

## IKEA's Knowledge Graph and Why It Has Three Layers

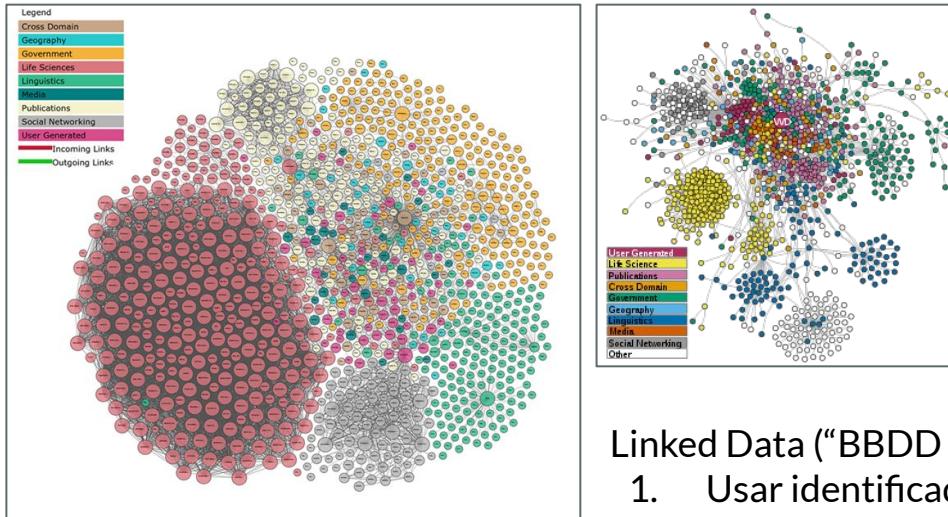
At IKEA we are building a knowledge graph to improve the overall experience of our customers in the physical and digital space. I like to use the following layered pyramid depiction when explaining to stakeholders what a Knowledge Graph (KG) consists of. I saw it for the first time when I watched [this video](#) by Dave McComb explaining the *gist ontology*.



# Knowledge Graphs - Wikidata



# Wikidata - Linked Open Data



Linked Data (“BBDD distribuida universal”):

1. Usar identificadores HTTP (URI) para conceptos
2. Cuando algo pide ese concepto, devolver datos en formatos abiertos
3. Incluir enlaces semánticos a otros conceptos

# Wikidata - Estándares interoperabilidad

RDF - Grafos (“HTML de los datos”)

SPARQL - Consultas sobre RDF (“SQL para grafos”)



URI - Identificar conceptos (“URLs para cosas”): Ej.

<http://www.wikidata.org/entity/Q8692>



# Wikidata - Estándares interoperabilidad

RDF - Grafos

SPARQL - Consultas

URI - Identificar conceptos y enlazar

Main page  
Community portal  
Project chat  
Create a new item  
Recent changes  
Random item  
[Query Service](#)  
Nearby  
Help  
Donate  
Lexicographical data  
Create a new Lexeme  
Recent changes  
Random Lexeme  
  
Tools  
What links here  
Related changes  
Special pages  
Permanent link  
Page Information  
Concept URI  
Cite this page

Item - Discussion

## Bilbao (Q8692)

municipality and city in Spain

In more languages  
Configure

Language	Label	Description
English	Bilbao	municipality and city in Spain
Spanish	Bilbao	ciudad de España, capital de la provincia española de Vizcaya
Catalan	Bilbao	ciutat del País Basc
Galician	Bilbao	cidade do País Vasco

All entered languages

Statements

Instance of: municipality of Spain (0 references)  
city

```
[mtkell@burruel BilbaoDataLab-Grafos]$ curl -L -H "Accept: application/ld+json" http://www.wikidata.org/entity/Q8692
{
  "@graph": [
    {
      "@id": "data/Q8692",
      "@type": "SchemaDataset",
      "about": "wd:Q8692",
      "license": "http://creativecommons.org/publicdomain/zero/1.0/",
      "softwareVersion": "1.0.0",
      "version": 1869492855,
      "dateModified": "2023-04-06T10:20:23Z",
      "statements": 162,
      "sitelinks": 142,
      "identifiers": 45
    },
    {
      "@id": "wd:Q8692",
      "@type": "WikibaseItem"
    },
    {
      "uri": "https://es.wikivoyage.org/wikivoyage/Bilbao",
      "@type": "SchemaArticle",
      "about": "wd:Q8692",
      "inLanguage": "es",
      "isPartOf": "https://es.wikivoyage.org/",
      "name": {
        "@language": "es",
        "@value": "Bilbao"
      }
    }
  ]
}
```

Wikidata Query Service Examples Help ▾

	1 <b>SELECT ?p ?o</b>
	2 <b>WHERE {</b>
	3    <http://www.wikidata.org/entity/Q8692> ?p ?o
	4 }

# Wikidata - Estándares interoperabilidad

1. Usar identificadores HTTP (URI) para conceptos: <http://www.wikidata.org/entity/Q8692>
2. Cuando algo pide ese concepto, devolver datos en formatos abiertos: [RDF](#), [SPARQL](#)
3. Incluir enlaces semánticos (Part-of, Works-for) a otros conceptos y así combinar datos:

```
<wdtn:P1566 rdf:resource="http://sws.geonames.org/6362368/" />
```

# Ejemplo de proceso

1. Añadir conjuntos de datos y metadatos
2. Convertir a RDF y reconciliar entidades/enlaces
3. Almacenaje
4. Publicación



# 1) Añadir conjuntos de datos y metadatos

The diagram illustrates the process of adding datasets and metadata to a GitHub repository, specifically the `BasqueCountryInstitutionsTransparentRelationsGraph`. A blue arrow points from the `People` dataset page to the GitHub repository, and another arrow points from the `Registro de licitadores y empresas clasificadas` dataset page to the same repository.

**BasqueCountryInstitutionsTransparentRelationsGraph** (Public)

- Pull requests: 0 | Discussions: 0 | Actions: 0 | Projects: 0 | Wiki: 0 | Security: 0 | Insights: 0 | Settings: 0
- Go to file | Add file | Code
- mikel-egana-arranguren No profile · 3 weeks ago · 17 commits
- People · No profile
- RegistroLicitadores\EmpresasClasif... · Improve README and add People example · 3 weeks ago · last month

**HORDAGO**

Boletines Tienda **Datos**

CIENCIA EGIPTO PANTERAS NEGRAS LEY DE VIVIENDA SECCIONES MEDIA EL SALTO ZONA SOCIALS

**PNV**

**La familia Atutxa (PNV): empresas y contratos públicos de Asier e Iskander**

Los hijos de Juan Muri Gurucea, Asier e Iskander han creado múltiples sociedades y obtenido millones de euros en contratos de la administración pública. Además, estas empresas han estado ligadas a casos de irregularidades ligadas al clientelismo en ayuntamientos

**People**

**Dataset**

- Name: People
- Source: Various sources, collected manually
- Metadata: [People.ttl](#) (Produced manually)

**Convert to RDF Graph**

- URI pattern:
  - Graph URI: <https://data.ehu.eus/bctr/people/graph>
  - Base URI: <https://data.ehu.eus/bctr/>
  - Person: <https://data.ehu.eus/bctr/person/>[normalised-person-name]
  - Role: <https://data.ehu.eus/bctr/role/>[normalised-person-name][officer|president|director]...][dates]
- Vocabularies used:
  - [Transparent Relations Ontology](#)
- `python3 CSV2RDF.py` (`pip install rdflib`)

**Store Graph**

- Upload [People.nq](#) (Data)
- Upload [People.ttl](#) (Metadata)

**Kontratacio Publikoa Euskadin**  
**plataforma**  
Contratación Pública en Euskadi

**Registro de licitadores y empresas clasificadas**

**Dataset**

- Name: Registro de licitadores y empresas clasificadas
- Source:
  - <https://www.contratacion.euskadi.eus/webkpe00-kperoc/es/k60aSolicitudesWar/InicioBusqEmpresas.do>
  - "Busqueda de empresas", "Informes": [Empresas-2023-03-05.pdf](#)
  - <https://www.adobe.com/acrobat/online/pdf-to-excel.html>: [Empresas-2023-03-05.csv](#)
- Metadata: [Empresas-2023-03-05-metadata.ttl](#) (Produced manually)

**Convert to RDF Graph**

- URI pattern:
  - Graph URI: <https://data.ehu.eus/bctr/graph>
  - Base URI: <https://data.ehu.eus/bctr/>
  - Company: <https://data.ehu.eus/bctr/company/>[Número inscripción]
- Vocabularies used:
  - [Transparent Relations Ontology](#)
- `python3 CSV2RDF.py` (`pip install rdflib`)

**Store Graph**

- Upload [Empresas-2023-03-05.nq](#) (Data)
- Upload [Empresas-2023-03-05-metadata.ttl](#) (Metadata)

## 2) Convertir a RDF y reconciliar entidades/enlaces



I HORDAGO Boletines Tienda **Data**

CHINIA EGIPTO PANTERAS NEGRAS LEY DE VIVIENDA SECCIONES MEDIA EL SALTO ZONA SOCIAL

**PNV**

**La familia Atutxa (PNV): empresas y contratos públicos de Asier e Iñaki**

Los hijos de Juan Mari Atutxa, Asier e Iñaki han creado múltiples sociedades y obtenido millones de contratos de la administración pública. Además, estas empresas han estado ligadas a casos de irregularidades ligadas al clientelismo en ayuntamientos vizcaínos

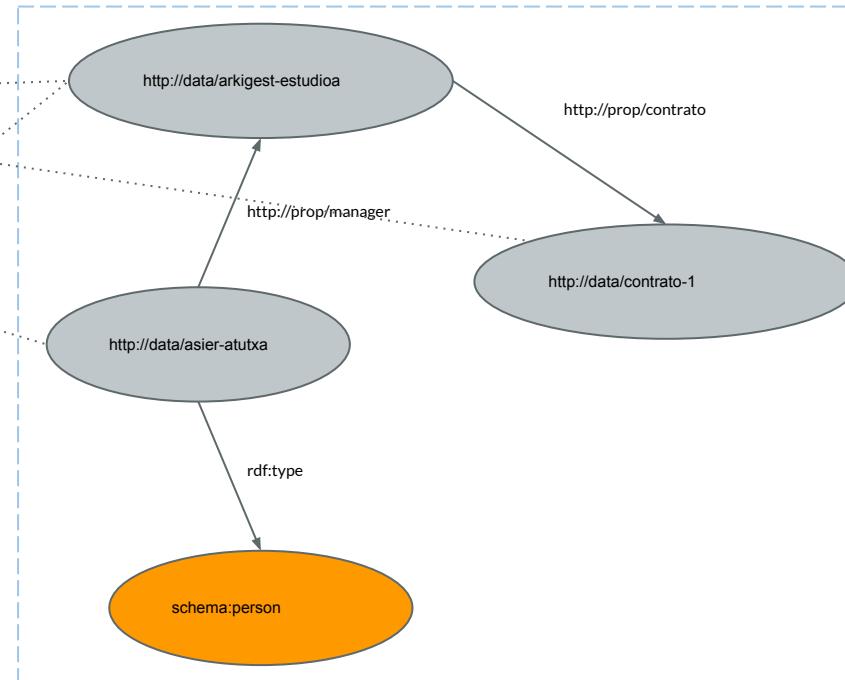
“ARKIGEST ESTUDIOA S.L.”

“ARKIGEST ESTUDIOA S.L.”

“Asier Atutxa”



```
for row in input_file:  
    if 6 < v < len(row.items()):  
        name = row['name']  
        profile = row['profile']  
        entity = row['entity']  
        role = row['role']  
        startdate = row['start-date']  
        enddate = row['end-date']  
        entity = row['entity']  
        evidence = row['evidence']  
  
# Código repetido de Registro Licitadores: mila ideas  
for row in input_file_licitadores:
```



```
<https://data.ehu.eus/bcltr/person/asier-atutxa-zaldunigis> <http://ehu.eus/triples/rule> :b  
<https://data.ehu.eus/bcltr/rele/asier-atutxa-zaldunigis:manager...arkigest.estudioa-s-l> :b  
<https://data.ehu.eus/bcltr/rf8d0/a/...asier-atutxa-zaldunigis> <http://www.w3.org/2002/07/owl/rdf:  
sameAs> <https://data.ehu.eus/bcltr/person/asier-atutxa-zaldunigis>  
<https://data.ehu.eus/bcltr/person/asier-atutxa-zaldunigis> <http://ehu.eus/triple/profile url>
```

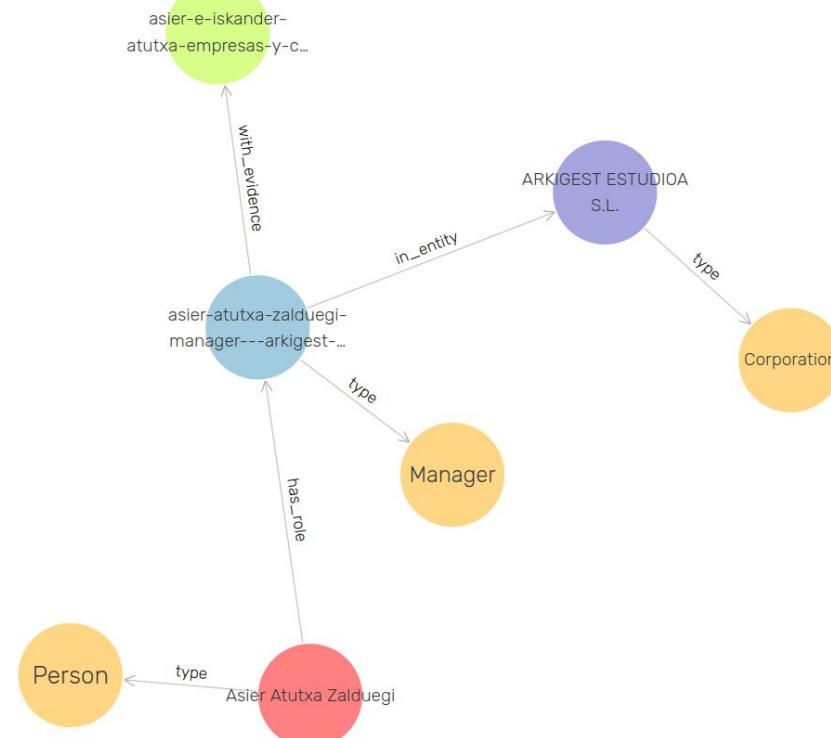
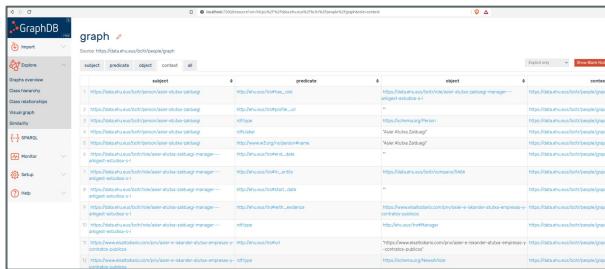
# 3) Almacenaje



```
<https://data.ehu.es/bc11r/person/asier-atutxa-zalduegi> <http://ehu.eus/tro#has_roles> <http://ehu.eus/tro#has_roles>
<https://data.ehu.es/bc11r/role/asier-atutxa-zalduegi-manager> , <http://www.w3.org/2000/01/rdf-schema#label> "asier-atutxa-zalduegi"
<https://data.ehu.es/bc11r/person/asier-atutxa-zalduegi> <http://www.w3.org/2000/01/rdf-schema#type> <http://ehu.eus/tro#profile_uris>
```



# Triple Store (GraphDB local)



# Triple Store (Blazegraph local)

WELCOME    QUERY    UPDATE    EXPLORE    NAMESPACES    STATUS    PERFORMANCE

Namespace shortcuts: Bigdata ▾ W3C ▾ Dublin Core ▾ SocialGraph ▾

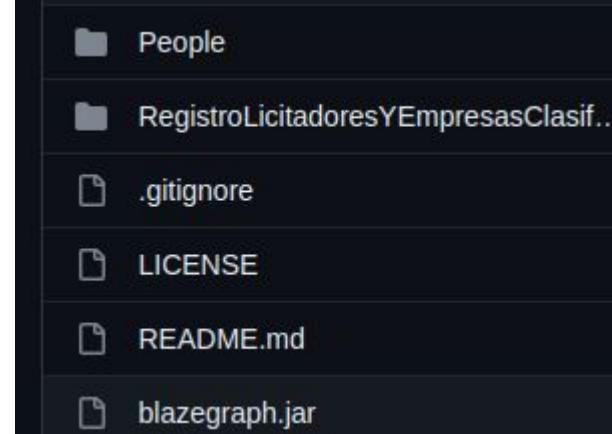
Web - SPARQL Query

```
1 SELECT *
2 WHERE {
3   GRAPH ?g {
4     ?s ?p ?o
5   }
6 }
```

Advanced features

Execute | Clear

S	P	O	S
<a href="#">http://data.ehu.eus/bcltr/person/aisier-atutxa-zalduegi#</a>	<a href="#">http://ehu.eus/trothas_role</a>	<a href="#">http://data.ehu.eus/bcltr/role/aisier-atutxa-zalduegi_manager...arkiest-estudioa-1.1#</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/person/aisier-atutxa-zalduegi#</a>	<a href="#">http://ehu.eus/trofersfile_url#</a>	<a href="#">http://www.v3.org/ns/person#name</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/person/aisier-atutxa-zalduegi#</a>	<a href="#">rdfs:type</a>	<a href="#">http://schema.org/Person</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/person/aisier-atutxa-zalduegi#</a>	<a href="#">rdfs:label</a>	<a href="#">Aisier Atutxa Zalduegi</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/sle/aisier-atutxa-zalduegi_manager...arkiest-estudioa-1.1#</a>	<a href="#">http://ehu.eus/trofend_date</a>	<a href="#">http://ehu.eus/trofend_date</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/sle/aisier-atutxa-zalduegi_manager...arkiest-estudioa-1.1#</a>	<a href="#">http://ehu.eus/trofin_entity</a>	<a href="#">http://data.ehu.eus/bcltr/company/5466#</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/sle/aisier-atutxa-zalduegi_manager...arkiest-estudioa-1.1#</a>	<a href="#">http://ehu.eus/trofstart_date</a>	<a href="#">http://ehu.eus/trofstart_date</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/sle/aisier-atutxa-zalduegi_manager...arkiest-estudioa-1.1#</a>	<a href="#">http://ehu.eus/trowith_evidence</a>	<a href="#">http://www.elsalvatiario.com/pms/aisier-e-iskander-atutxa-empresas-y-contratos-publicos#</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://data.ehu.eus/bcltr/sle/aisier-atutxa-zalduegi_manager...arkiest-estudioa-1.1#</a>	<a href="#">rdf:type</a>	<a href="#">http://ehu.eus/trofManager</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://www.elsalvatiario.com/pms/aisier-e-iskander-atutxa-empresas-y-contratos-publicos#</a>	<a href="#">http://ehu.eus/trofUrl</a>	<a href="#">https://www.elsalvatiario.com/pms/aisier-e-iskander-atutxa-empresas-y-contratos-publicos#</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>
<a href="#">http://www.elsalvatiario.com/pms/aisier-e-iskander-atutxa-empresas-y-contratos-publicos#</a>	<a href="#">rdfs:label</a>	<a href="#">http://schema.org/NewArticle</a>	<a href="#">https://data.ehu.eus/bcltr/people/raphael-zalduegi#</a>



# 4) Publicación



A screenshot of a SPARQL query interface at [api.euskadi.eus/sparql/](http://api.euskadi.eus/sparql/). The query is:

```
1 ▾ SELECT * WHERE {  
2   ?sub ?pred ?obj .  
3 }  
4 LIMIT 10
```

The results pane shows a single row of data:

type	Person
has_role	asier-atutxa-zalduegi-manager—arkigest-estudioas-s-i
profile_url	
label	Asier Atutxa Zalduegi
name	Asier Atutxa Zalduegi

Below the results, there are buttons for 'Raw Response', 'Table', 'Pivot Table', 'Google Chart', and a download icon.



## Asier Atutxa Zalduegi

<http://localhost:8080/bcitr/person/asier-atutxa-zalduegi>

a <https://schema.org/Person>

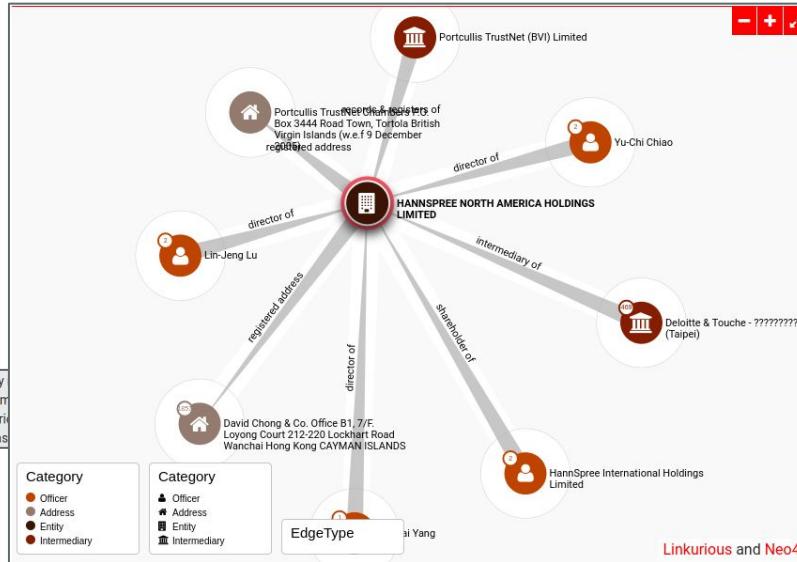
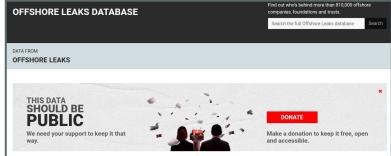
type	Person
has_role	asier-atutxa-zalduegi-manager—arkigest-estudioas-s-i
profile_url	
label	Asier Atutxa Zalduegi
name	Asier Atutxa Zalduegi

[json-id](#) | [turtle](#) | [n3](#)



# Otros proyectos

# Offshore Leaks Database



"Papeles de Panamá"

- (+) Gran trabajo colectivo de investigación e integración
- (+) Datos publicados
- (+) Buen interfaz
- (-) Neo4J (No estándar)
- (-) Silo de datos: no publican URIs (No enlazable)

# La donación



## Donaciones al rey emérito

- (-) Individual
- (-) Pila tecnológica completa desde cero (WTF! - No estándar)
- (-) Silo de datos: no publican URIs (No enlazable)
- (-) No publica datos brutos
- (+) Buen interfaz

No todo el monte es orégano

# There is no free lunch

Integrar información es muy difícil, independientemente de la tecnología elegida:

- Identificadores
- Reconciliación de entidades
- Modelado de datos
- Interoperabilidad
- ...

El trabajo lo vamos a tener que hacer igual-igual, pero los KGs:

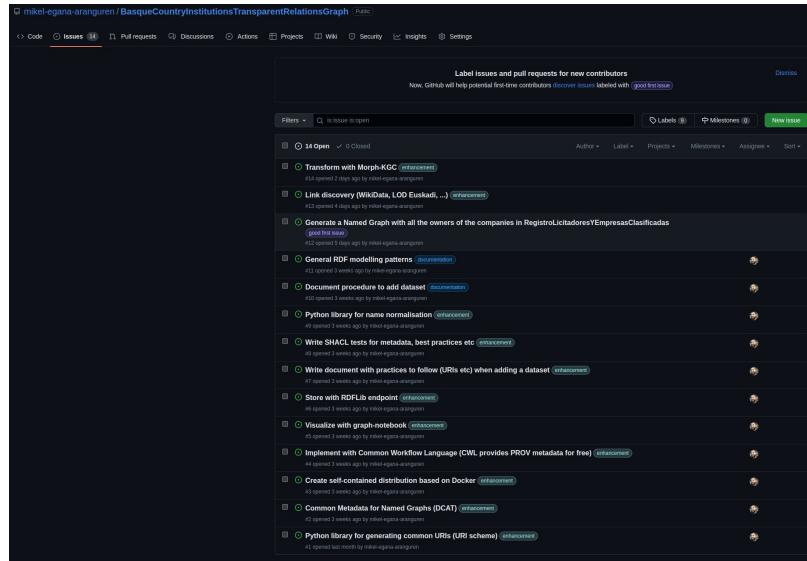
- Nos permiten centrarnos en los problemas de integración
- Nos permiten reusar el trabajo de integración

# Posibles tareas y grupos de trabajo

# Generate a Named Graph with all the owners of the companies in RegistroLicitadoresYEmpresasClasificadas

good first issue

#12 opened last month by mikel-egana-aranguren



mikel-egana-aranguren / BasqueCountryInstitutionsTransparentRelationsGraph · [Code](#) · [Issues](#) · [Pull requests](#) · [Discussions](#) · [Actions](#) · [Projects](#) · [Wiki](#) · [Security](#) · [Insights](#) · [Settings](#)

Label issues and pull requests for new contributors  
Now, GitHub will help potential first-time contributors discover issues labeled with `good first issue`

Filters · [Issues to open](#) · [Labels](#) (0) · [Milestones](#) (0) · [New issue](#)

14 Open · 0 Closed

Author · Label · Projects · Milestones · Assignee · Sort

14 · [Transform with Morph-KGC](#) ([enhancement](#)) · [Link discovery \(Wikidata, LOD Euskadi, ...\)](#) ([enhancement](#)) · [Generate a Named Graph with all the owners of the companies in RegistroLicitadoresYEmpresasClasificadas](#) ([good first issue](#)) · [General RDF modeling patterns](#) ([enhancement](#)) · [Document procedure to add dataset](#) ([enhancement](#)) · [Python library for name normalisation](#) ([enhancement](#)) · [Write SHACL tests for metadata, best practices etc](#) ([enhancement](#)) · [Write document with practices to follow \(URIs etc\) when adding a dataset](#) ([enhancement](#)) · [Store with RDFLib endpoint](#) ([enhancement](#)) · [Visualize with graph-notebook](#) ([enhancement](#)) · [Implement with Common Workflow Language \(CWL provides PROV metadata for free\)](#) ([enhancement](#)) · [Create self-contained distribution based on Docker](#) ([enhancement](#)) · [Common Metadata for Numei Graphs \(DCAT\)](#) ([enhancement](#)) · [Python library for generating common URIs \(URI scheme\)](#) ([enhancement](#))



## Perfiles:

- Técnico
- Documentación
- Análisis

## Debate / Trabajo en grupo

Organización / Legal

Ejercicios

Consumo de datos

Investigación

Catálogo datasets

Buenas prácticas

Desarrollo, almacenaje  
y publicación

mikel egaña aranguren github

All Images Videos News Maps Settings

Spain (ca) Safe search: moderate Any time

<https://github.com/mikel-egana-aranguren>

**mikel-egana-aranguren (Mikel Egaña Aranguren) · GitHub**

Mikel Egaña Aranguren, Ph.D. Background My research and development activity is related to the publication and consumption of data following FAIR principles (Findable, Accessible, Interoperable, Reusable), mainly in Government Open Data and Life Sciences, usually (but not only) through Linked Data.

Overview Repositories 109 Projects Packages Stars 516

Basque Type Language Sort New

1 result for repositories matching Basque sorted by last updated Clear filter

**BasqueCountryInstitutionsTransparentRelationsGraph** Public

Transparent relations involving Basque Country institutions

Star

Python Apache License 2.0 1 issue needs help Updated yesterday

Pull requests Discussions Actions Projects Wiki Security

bdl 2 branches 0 tags

This branch is 8 commits ahead of main.

mikel-egana-aranguren consulta

EjercicioTransformacionRDF EjercicioVisualizacionRDF

# Organización

¿El proyecto sigue más allá de mi “side-project”? ¿Como?

¿Alguien más participa? ¿Cómo?

Gestión del desarrollo (GitFlow, Kanban, GitHub projects, ...)

Repositorio del proyecto (¡No en el mío personal!)

Servidores/Recursos colectivos

...

# Legal

¿Qué datos se pueden reutilizar? (Ej. ¿No se puede “escapar” LinkedIn pero se puede reutilizar el registro de licitadores?)

¿Qué datos se van a hacer públicos?

¿Se puede integrar información de una persona concreta de diferentes fuentes, aunque estas fuentes sean públicas y reutilizables (Ej. Open Data, noticias, ...)?

¿Dónde se van a hacer públicos?

¿Qué consecuencias tiene publicar la información, aunque sea “Publicable”?

Licencias

etc.

# Catálogo datasets

No hace falta conocimiento técnico

Buscar datasets interesantes

Crear una ficha y metadatos

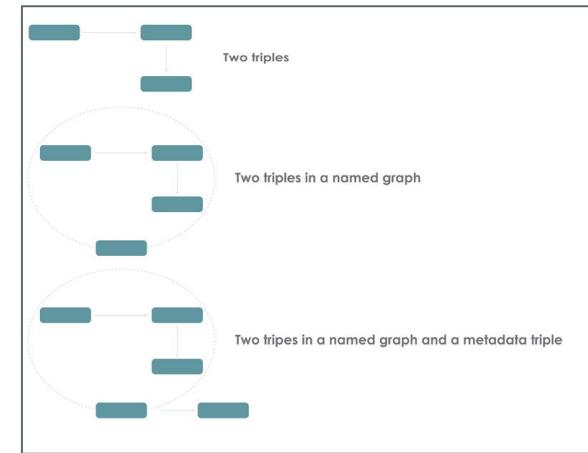
Crear un mapeo sobre un dataset ya existente (ej.  
MySQL contratos)

Se puede usar otra tecnología a partir del catálogo,  
no sólo grafos

**Registro de licitadores y empresas clasificadas**

**Dataset**

- Name: Registro de licitadores y empresas clasificadas
- Source:
  - <https://www.contratacion.euskadi.eus/webkpe00-kperoc/es/k60aSolicitudesWar/InicioBusqEmpresas.do>
  - "Busqueda de empresas", "Informes": [Empresas-2023-03-05.pdf](#)
  - <https://www.adobe.com/acrobat/online/pdf-to-excel.html>: [Empresas-2023-03-05.csv](#)
- Metadata: [Empresas-2023-03-05-metadata.ttl](#) (Produced manually)



# Buenas prácticas

## Esquema para URLs:

- Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información (NTI)
- European Legislation Identifier (ELI)
- ...

### *URI para identificar a cualquier instancia física o conceptual*

Estos recursos son las representaciones atómicas de los documentos y recursos de información. A su vez suelen ser instancias de las clases que se definen en los vocabularios. Estos recursos se identifican mediante el esquema: <http://{base}/recurso/{sector}/{dominio}/{clase}/{ID}>



## Metadatos:

- Esquema general
- DCAT, VOID, PROV, ...
- Validación mediante SHACL

## Formato de fichas de datasets

## Documentación

### 1. *URI template/patterns*

ELI's HTTP URLs enable users to access legislation in a persistent way. The URLs are formally described by machine-readable templates, using semantic components from a legal and an end-user point of view.

# Desarrollo, almacenaje y publicación

Triple Store, Linked Data, APIs

Integración continua (GitHub actions)

Despliegue, DevOps, ...

Reconciliación de entidades (Silk, ...)

NLP, NER, Text mining, ...

Creación y documentación de módulos Python comunes: normalización nombres, esquema URLs, ...

# Consumo de datos

Algoritmos (Análisis de redes, sistemas de recomendación, ...)

SPARQL + R

Visualizaciones (SPARQL + D3, AWS Graph Jupyter Notebook, ...)

Aplicaciones web

Consultas federadas mediante SPARQL

Linked Data: negociación de contenido

...

# Universidad

Possible tesis doctoral y/o TFG/TFM : “Plataforma basada en grafos de conocimiento para la representación y explotación de datos sobre posibles conflictos de interés”

- Plataforma con interfaz gráfico para aumentar usabilidad
- Arquitectura con un mínimo de calidad
- Despliegue como recurso público, con URIs persistentes ([w3id.org](http://w3id.org))
- Funciones adicionales: Natural Language Processing (NLP), Named Entity Recognition (NER), reconciliación avanzada de entidades, ChatGPT, ...

# Universidad

mikel-egana-aranguren / Transparent-Relations-Ontology Public

< Code Issues Pull requests Discussions Actions Projects Wiki Security Insights Settings

Code develop 7 branches 0 tags Go to file Add file <> Code

mikel-egana-aranguren Add PPROC as related project 1489938 2 weeks ago 28 commits

File	Commit Message	Time
github/workflows	Move ROBOT stuff to robot/	last year
development	Correct version IRI	last month
release	Improve annotations	last month
robot	Widoco release	last year
.gitignore	Move ROBOT stuff to robot/	last year
LICENSE	First commit	last year
README.md	Add PPROC as related project	2 weeks ago
RELEASES.md	Release 0.1.1	last month
widoco.sh	Add proper .htaccess	last year

README.md

License Apache2.0 Makefile CI pending

Transparent Relations Ontology



Ejercicios  
Preguntas  
Debate / Trabajo en grupo

