

**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, ...

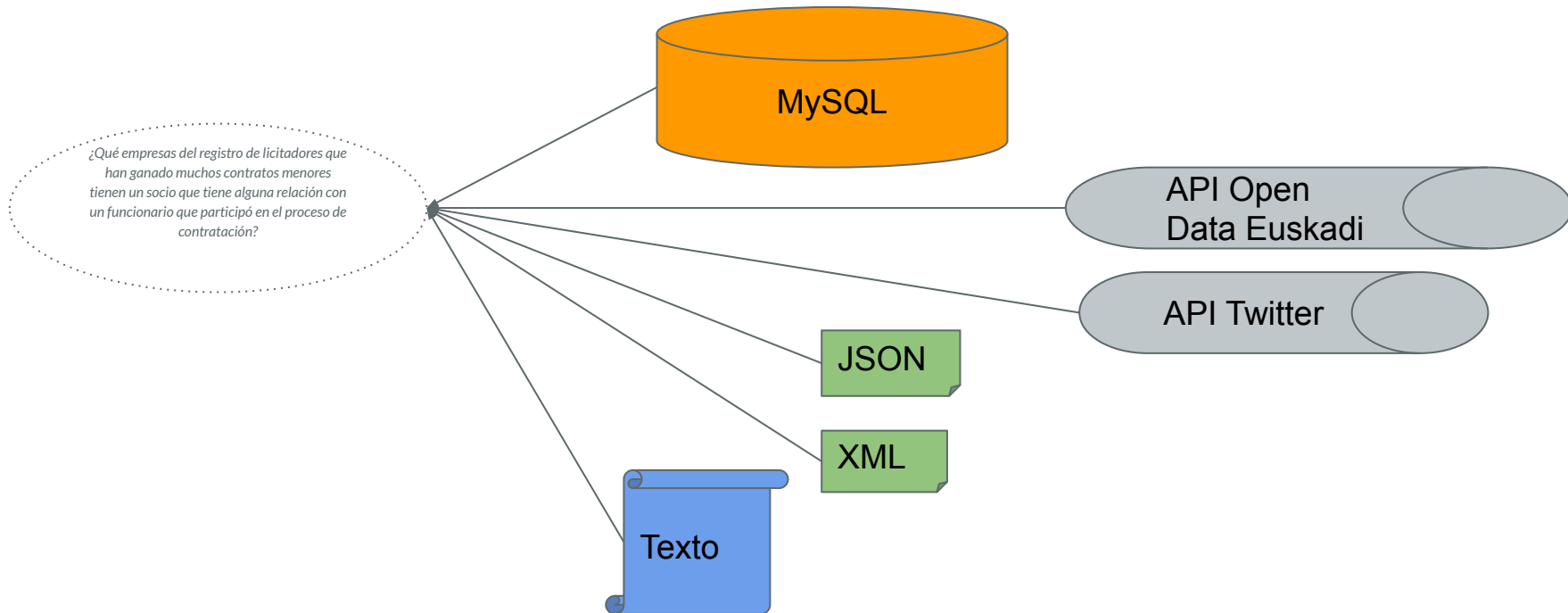
# Integrar

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

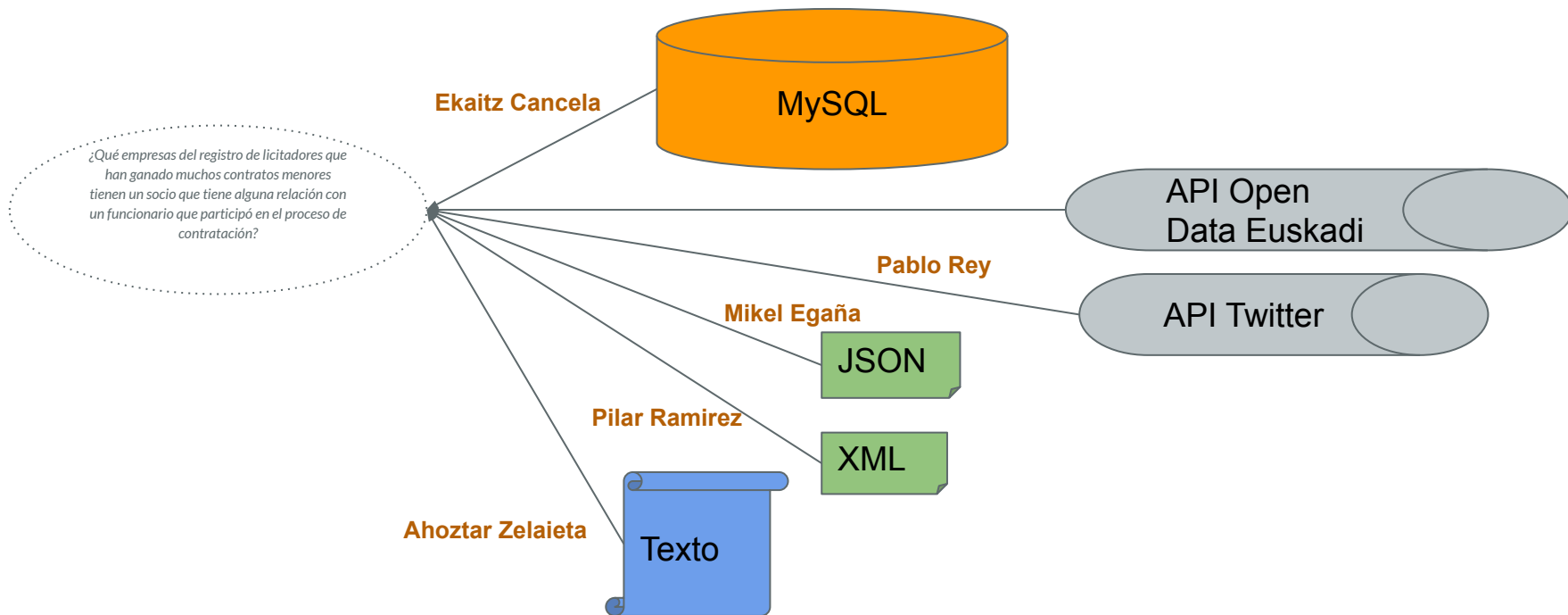
*¿Qué empresas del registro de licitadores que han ganado muchos contratos menores tienen un socio que tiene alguna relación con un funcionario que participó en el proceso de contratación?*



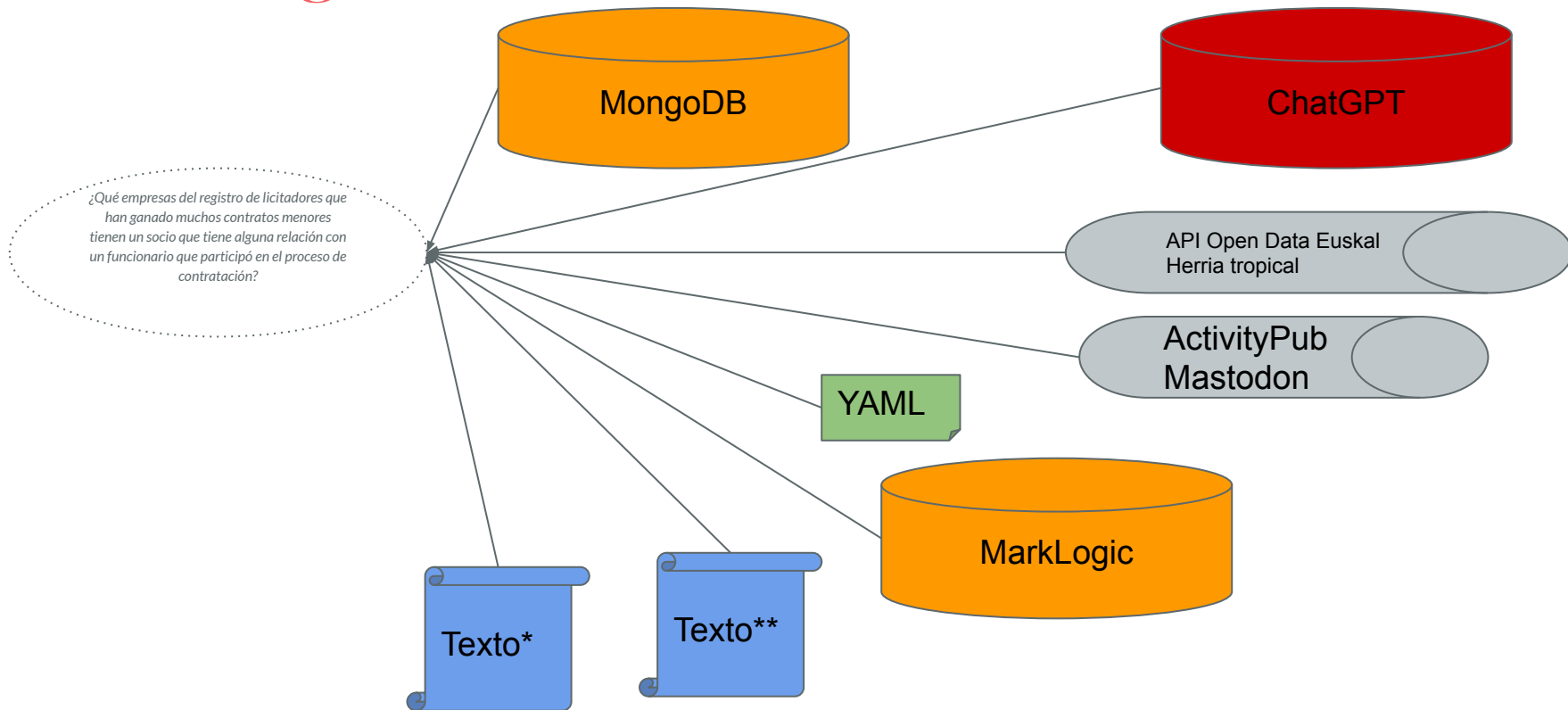
# Integrar



# De manera colaborativa

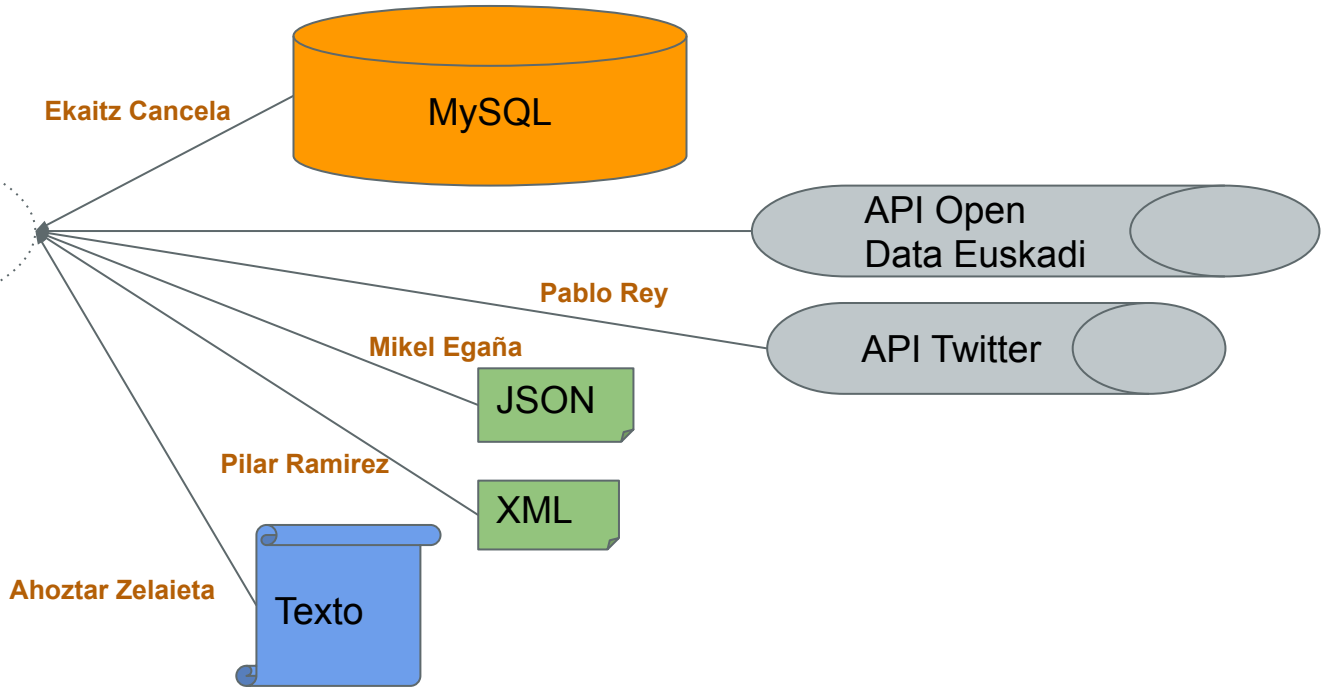


# Tecnológicamente sostenible

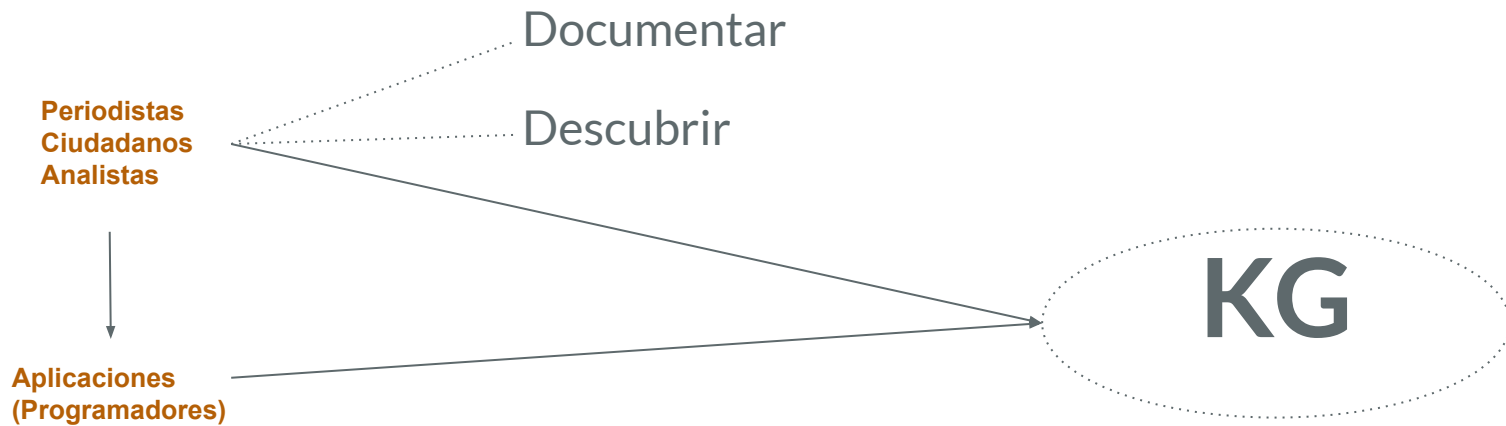


# KG

Integrar  
De manera colaborativa  
Tecnológicamente sostenible

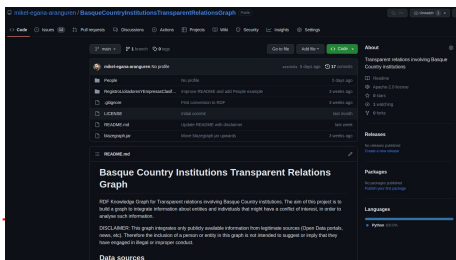


# Publicar



The screenshot shows a GitHub repository page for 'BasqueCountryInstitutionsTransparentRelationsGraph'. The repository is public and has 14 issues, 1 pull request, and 17 commits. The main branch is 'main'. The repository contains several files: 'People', 'RegistroLicitadoresYEmpresasClasif...', '.gitignore', 'LICENSE', 'README.md', and 'blazegraph.jar'. The README.md file is open, showing the title 'Basque Country Institutions Transparent Relations Graph' and a description: '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.' A disclaimer follows, stating that the graph integrates only publicly available information from legitimate sources. The 'Data sources' section is also visible. The right sidebar shows the 'About' section, which includes the repository name, a README link, an Apache-2.0 license, 0 stars, 1 watching, and 0 forks. The 'Releases' section shows no releases published. The 'Packages' section shows no packages published. The 'Languages' section shows Python at 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

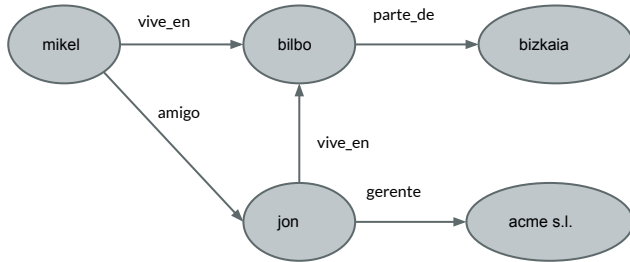
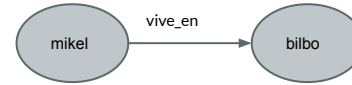


# Knowledge Graphs

**Graph:** Modelo de datos basado en nodos y vértices

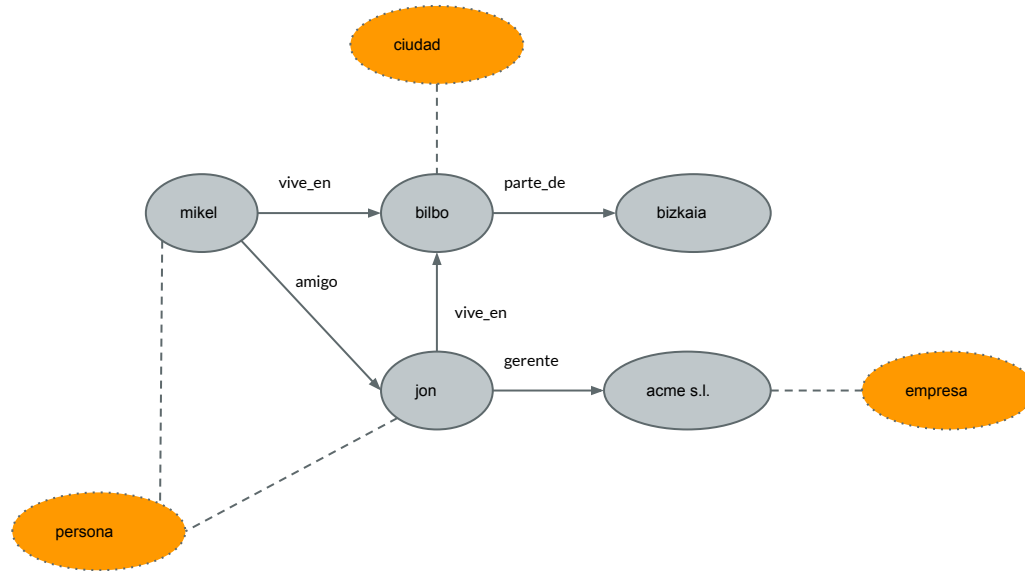
Tripleta formada por dos nodos y un vértice

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

**Schema.org**

Person

A Schema.org Type

Thing > Person

A person (alive, dead, undead, or fictional).

Property	Expected Type	Description
<b>Properties from Person</b>		
<b>Address</b>	Text	An address used for a person.
<b>Address</b>	PostalAddress or Text	Physical address of the person.
<b>Affiliation</b>	Organization	An organization that this person is affiliated with.
<b>AlumniOf</b>	EducationalOrganization or Organization	An organization that this person was a student of.



EXAMPLE

## A reintroduction to our Knowledge Graph and knowledge panels

May 10, 2020 · 2 min read

**Denny Soltan**  
Product Manager for Search

Sometimes Google Search will show special boxes with information about people, places and things. We call these knowledge panels. These designed to help you quickly understand items and in particular edges by surfacing key facts and to make it easier to explore topics in more depth. Information within knowledge panels comes from our Knowledge Graph, which is the giant web of relationships of facts. In this post, we'll share more about how knowledge panels are automatically generated from data in the Knowledge Graph to give you an idea how we use metadata to enrich or enhance information.

**What's a knowledge panel?**

Knowledge panels are easily recognized by those who do desktop searching, appearing to the right of search results.

### Enterprise Knowledge Graph overview

**Preview**

This product or feature is covered by the [EULA](#), [Privacy Policy](#), [Terms of Service](#). The GA products and features might have limited support, and changes to non-GA products and features might not be compatible with other non-GA features. For more information, see the [legal notice](#).

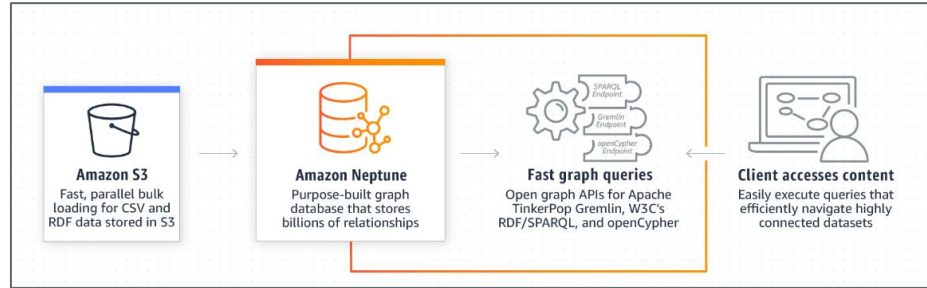
Enterprise Knowledge Graph organizes siloed information into organizational knowledge, which involves consolidating, standardizing, and reconciling data in an efficient and useful way.

#### Entity Reconciliation API

Entity Reconciliation API is a lightweight, AI-powered, clustering and deduplication service for tabular data. It is a standalone API that wraps around the Google open entity reconciliation engine. It helps customers to reconcile and join their data together, or to join their data with ours or multiple third party datasets.

**Key features include:**

- State-of-the-art quality optimized model with Google data.
- Batch entity enrichment, including geocoding.
- Google-scale clustering and reconciliation that handles a graph with a size up to billions of nodes and billions of edges.
- Stable Machine ID (SMID) is a unique identifier for each of the entity clusters.
- Native support up to 10 BigQuery tables.



# 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

Main Page Discussion

Read View source View history

collaborative linked structured

### Welcome to Wikidata

the free knowledge base with 102,734,855 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

Want to help translate? Translate the missing messages.

free open

- Main page
- Community portal
- Project chat
- Create a new item
- Recent changes
- Random item
- Query Service
- Nearby
- Help
- Donate

WIKIDATA

Item Discussion

## Bilbao (Q8692)

municipality and city in Spain

In more languages

### Statements

instance of

= 0 references

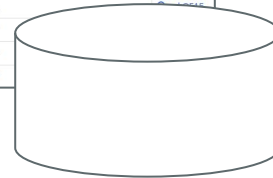
= 0 references

image

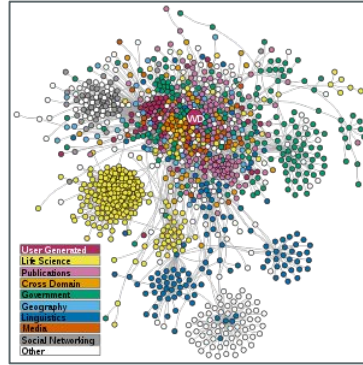
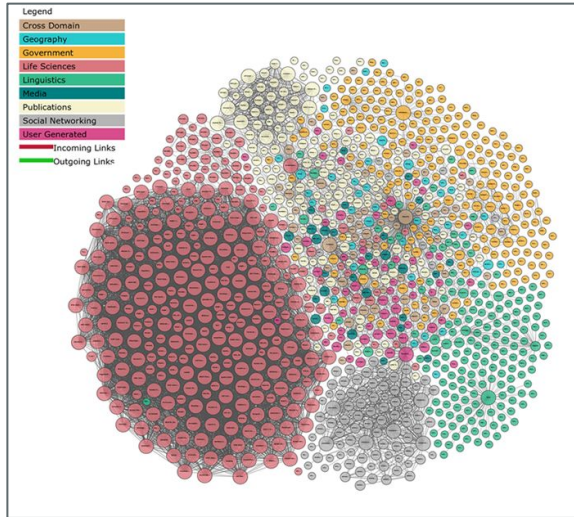
Cityscape - Bilbao, Spain - panoramio (1).jpg

bilbo

wd:Q8692	rdfs:label	Bilbao
wd:Q8692	rdfs:label	Bilbao
wd:Q8692	wd:P31	Leizor
wd:Q8692	wd:P37	
wd:Q8692	wd:P37	
wd:Q8692	wd:P17	



# 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

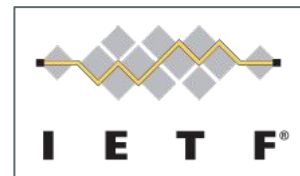
# Wikidata - Estándares interoperabilidad

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

```
[mikel@durru1 BilbaoDataLab-Grafos]$ curl -L -H "Accept: application/ld+json" http://www.wikidata.org/entity/Q8692
{
  "@graph": [
    {
      "@id": "data:Q8692",
      "@type": "schema:Dataset",
      "about": "wd:Q8692",
      "license": "https://creativecommons.org/publicdomain/zero/1.0/",
      "softwareVersion": "1.0.0",
      "version": "1869492055",
      "dateModified": "2023-04-06T10:20:23Z",
      "statements": 162,
      "siteLinks": 142,
      "identifiers": 45
    },
    {
      "@id": "wd:Q8692",
      "@type": "wikibase:Item"
    },
    {
      "@id": "https://es.wikivoyage.org/wiki/Bilbao",
      "@type": "schema:Article",
      "about": "wd:Q8692",
      "inLanguage": "es",
      "isPartOf": "https://es.wikivoyage.org/",
      "name": {
        "@language": "es",
        "@value": "Bilbao"
      }
    }
  ]
}
```

Item Discussion

## Bilbao (Q8692)

municipality and city in Spain

▼ In more languages

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	<span><span><span></span></span></span> municipality of Spain
	▼ 0 references
	<span><span><span></span></span></span> city

Wikidata Query Service

Examples Help

```
1 SELECT ?p ?o
2 WHERE {
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:

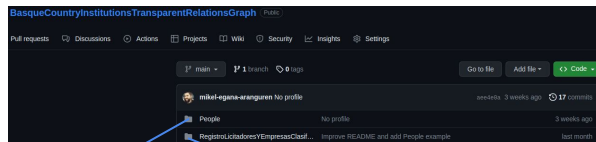
```
<wdtn:P1566 rdf:resource="http://sws.geonames.org/6362368/" />
```

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



## 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/bcitr/people/graph`
  - Base URI: `https://data.ehu.eus/bcitr/`
  - Person: `https://data.ehu.eus/bcitr/person/[normalised-person-name]`
  - Role: `https://data.ehu.eus/bcitr/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)

## Registro de licitadores y empresas clasificadas

### Dataset

- Name: Registro de licitadores y empresas clasificadas
- Source:
  - `https://www.contratacion.euskadi.eus/webkpe00-kperoc/es/60aSolicitudesWar/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)

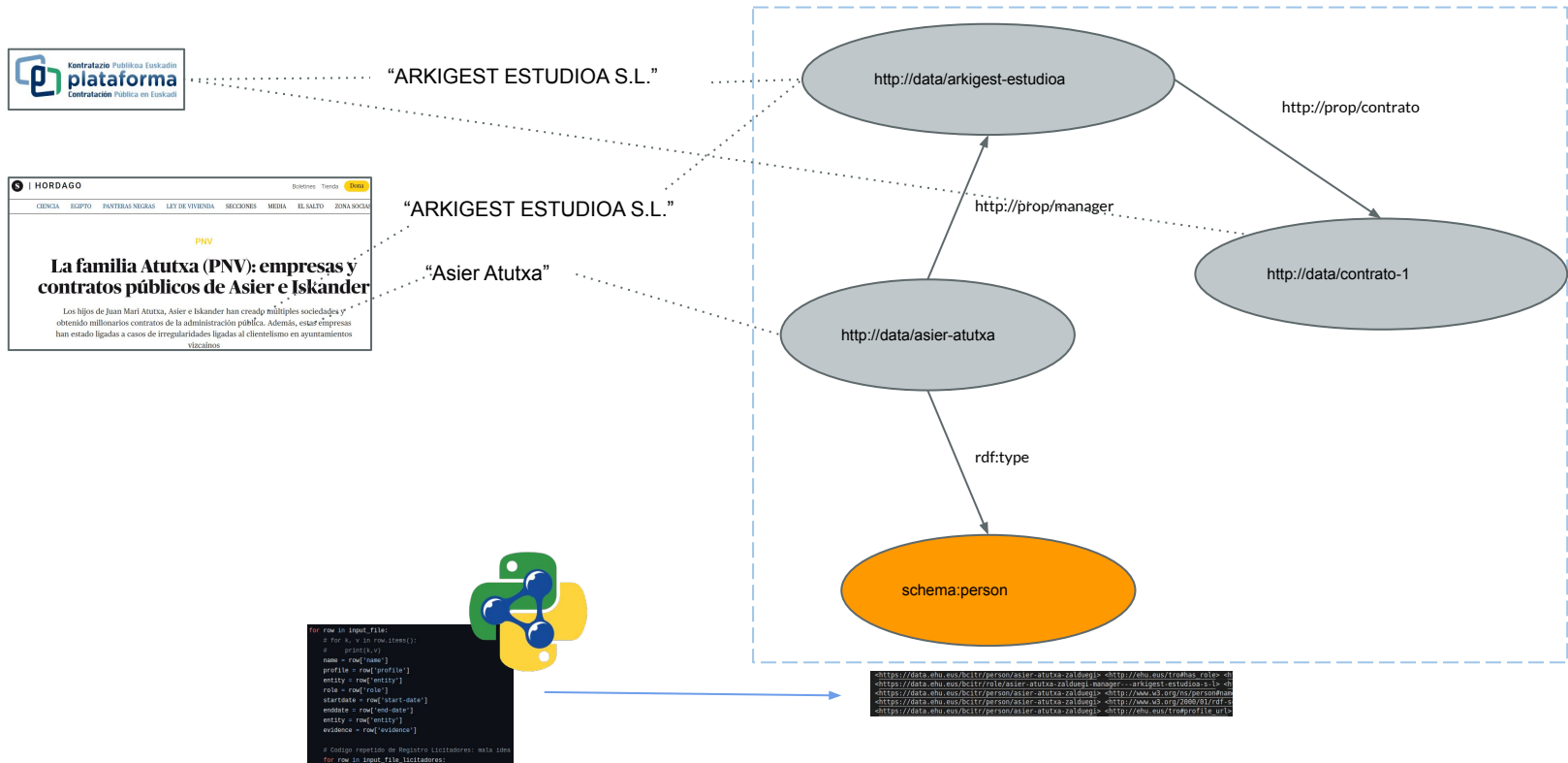
### Convert to RDF Graph

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

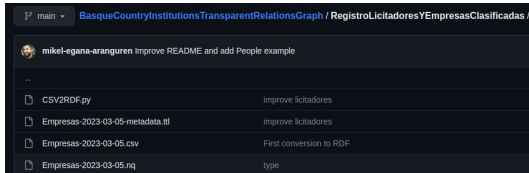
### Store Graph

- Upload `Empresas-2023-03-05.nq` (Data)
- Upload `Empresas-2023-03-05-metadata.ttl` (Metadata)

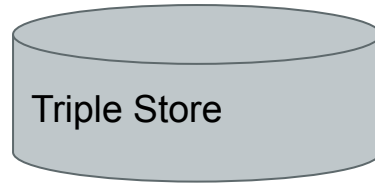
# 2) Convertir a RDF y reconciliar entidades/enlaces



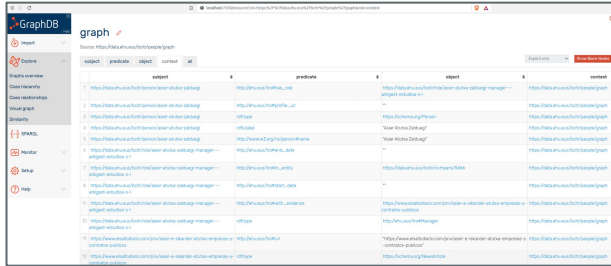
# 3) Almacenaje



```
<https://data.ehu.es/bc11r/person/asier_atutxa_zaldugi> <http://ehu.es/110m8s_role>  
<https://data.ehu.es/bc11r/role/asier_atutxa_zaldugi_manager> <arkivet-estudioa-s13>  
<https://data.ehu.es/bc11r/person/asier_atutxa_zaldugi> <http://www.w3.org/ns/person#>  
<https://data.ehu.es/bc11r/person/asier_atutxa_zaldugi> <http://www.w3.org/2000/01/rdf-schema#>  
<https://data.ehu.es/bc11r/person/asier_atutxa_zaldugi> <http://ehu.es/110m8s_role>
```

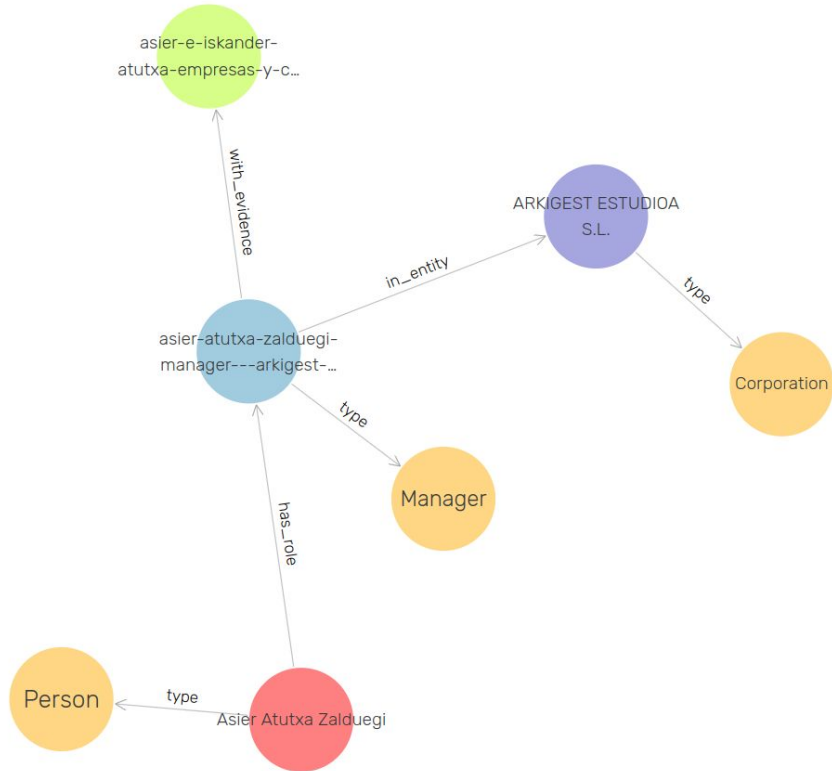


# Triple Store (GraphDB local)



The screenshot shows the GraphDB interface with a table of triples. The table has columns for subject, predicate, object, and evidence. The data is as follows:

subject	predicate	object	evidence
<a href="http://data.arkigest.com/ontology/arkigest#atutxa">http://data.arkigest.com/ontology/arkigest#atutxa</a>	<a href="http://data.arkigest.com/ontology/arkigest#with_evidence">http://data.arkigest.com/ontology/arkigest#with_evidence</a>	<a href="http://data.arkigest.com/ontology/arkigest#asier_e_iskander_atutxa_empresas_y_c...">http://data.arkigest.com/ontology/arkigest#asier_e_iskander_atutxa_empresas_y_c...</a>	<a href="http://data.arkigest.com/ontology/arkigest#with_evidence">http://data.arkigest.com/ontology/arkigest#with_evidence</a>
<a href="http://data.arkigest.com/ontology/arkigest#atutxa">http://data.arkigest.com/ontology/arkigest#atutxa</a>	<a href="http://data.arkigest.com/ontology/arkigest#in_entity">http://data.arkigest.com/ontology/arkigest#in_entity</a>	<a href="http://data.arkigest.com/ontology/arkigest#arkigest_estudioa_s_l">http://data.arkigest.com/ontology/arkigest#arkigest_estudioa_s_l</a>	<a href="http://data.arkigest.com/ontology/arkigest#in_entity">http://data.arkigest.com/ontology/arkigest#in_entity</a>
<a href="http://data.arkigest.com/ontology/arkigest#atutxa">http://data.arkigest.com/ontology/arkigest#atutxa</a>	<a href="http://data.arkigest.com/ontology/arkigest#type">http://data.arkigest.com/ontology/arkigest#type</a>	<a href="http://data.arkigest.com/ontology/arkigest#corporation">http://data.arkigest.com/ontology/arkigest#corporation</a>	<a href="http://data.arkigest.com/ontology/arkigest#type">http://data.arkigest.com/ontology/arkigest#type</a>
<a href="http://data.arkigest.com/ontology/arkigest#atutxa">http://data.arkigest.com/ontology/arkigest#atutxa</a>	<a href="http://data.arkigest.com/ontology/arkigest#type">http://data.arkigest.com/ontology/arkigest#type</a>	<a href="http://data.arkigest.com/ontology/arkigest#manager">http://data.arkigest.com/ontology/arkigest#manager</a>	<a href="http://data.arkigest.com/ontology/arkigest#type">http://data.arkigest.com/ontology/arkigest#type</a>
<a href="http://data.arkigest.com/ontology/arkigest#atutxa">http://data.arkigest.com/ontology/arkigest#atutxa</a>	<a href="http://data.arkigest.com/ontology/arkigest#has_role">http://data.arkigest.com/ontology/arkigest#has_role</a>	<a href="http://data.arkigest.com/ontology/arkigest#asier_atutxa_zalduegi">http://data.arkigest.com/ontology/arkigest#asier_atutxa_zalduegi</a>	<a href="http://data.arkigest.com/ontology/arkigest#has_role">http://data.arkigest.com/ontology/arkigest#has_role</a>
<a href="http://data.arkigest.com/ontology/arkigest#atutxa">http://data.arkigest.com/ontology/arkigest#atutxa</a>	<a href="http://data.arkigest.com/ontology/arkigest#type">http://data.arkigest.com/ontology/arkigest#type</a>	<a href="http://data.arkigest.com/ontology/arkigest#person">http://data.arkigest.com/ontology/arkigest#person</a>	<a href="http://data.arkigest.com/ontology/arkigest#type">http://data.arkigest.com/ontology/arkigest#type</a>





# Triple Store (Blazegraph local)

WELCOME QUERY UPDATE EXPLORE NAMESPACES STATUS PERFORMANCE

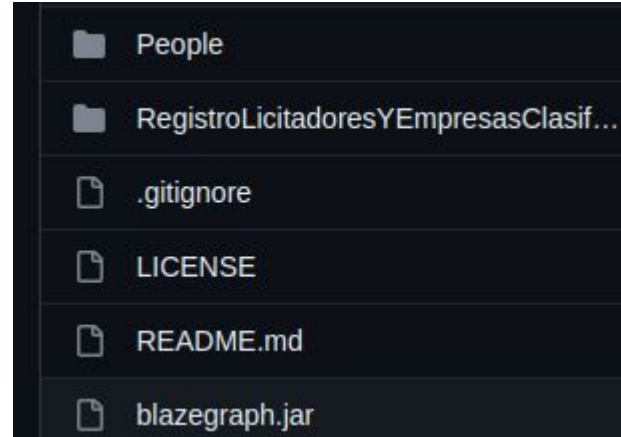
WQL - SPARQL Query Namespace shortcuts: [Bigdata](#) [W3C](#) [Dublin Core](#) [Social Ont](#)

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

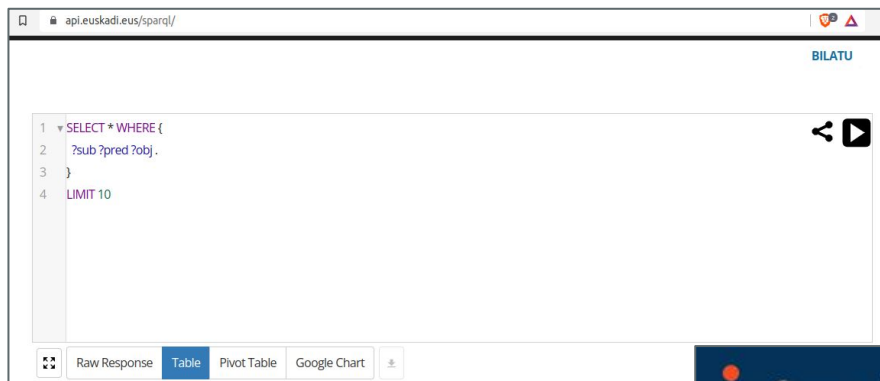
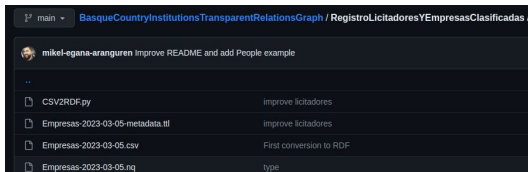
Advanced features

Execute Clear

s	p	o	g
<a href="http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi">http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="https://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l">https://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi">http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi">http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	Asier Atuxa Zalduegi	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi">http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="https://schema.org/Person">https://schema.org/Person</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi">http://data.ehu.es/blazegraph/person/asier-atuxa-zalduegi</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	Asier Atuxa Zalduegi	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l">http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l">http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="https://data.ehu.es/blazegraph/empresa/546">https://data.ehu.es/blazegraph/empresa/546</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l">http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l">http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="https://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos">https://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l">http://data.ehu.es/blazegraph/role/asier-atuxa-zalduegi-manager...arkitect-estudios-s-l</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos">http://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="https://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos">https://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>
<a href="http://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos">http://www.elalodiario.com/ov/asier-e-iskander-atuxa-empresas-y-contratos-publicos</a>	<a href="http://data.ehu.es/rofnas/role">http://data.ehu.es/rofnas/role</a>	<a href="https://schema.org/NewsArticle">https://schema.org/NewsArticle</a>	<a href="http://data.ehu.es/blazegraph/empresa">http://data.ehu.es/blazegraph/empresa</a>

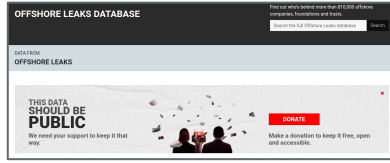


# 4) Publicación

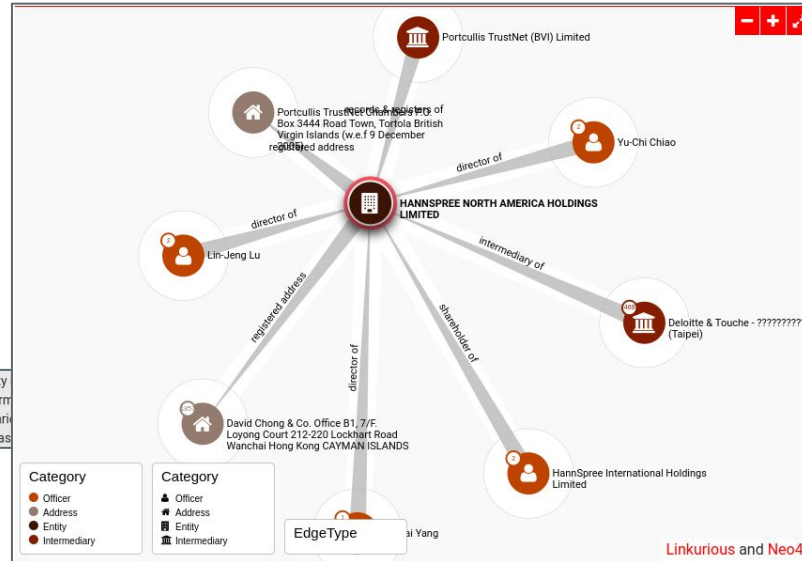


Otros proyectos

# Offshore Leaks Database



**DISCLAIMER:** There are legitimate uses for offshore companies and trusts. The inclusion of a person or entity illegal or improper conduct. Many people and entities have the same or similar names. We suggest you confirm identifiable information. The data comes directly from the leaked files ICLJ has received in connection with various databases. Some information may have changed over time. **Please contact us** if you find an error in the 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

The screenshot shows the GitHub interface for the repository 'BasqueCountryInstitutionsTransparentRelationsGraph'. The 'Issues' tab is active, displaying a list of 14 open issues. The issue 'Generate a Named Graph with all the owners of the companies in RegistroLicitadoresYEmpresasClasificadas' is highlighted in blue. It is labeled as a 'good first issue' and was opened 3 days ago by mikel-egana-aranguren. Other issues in the list include 'Transform with Morph-KGC', 'Link discovery (Wikidata, LOD Euskadi, ...)', 'General RDF modelling patterns', 'Document procedure to add dataset', 'Python library for name normalisation', 'Write SHACL tests for metadata, best practices etc', 'Write document with practices to follow (URIs etc) when adding a dataset', 'Store with RDFLib endpoint', 'Visualize with graph-notebook', 'Implement with Common Workflow Language (CWL, provides PROV metadata for free)', 'Create self-contained distribution based on Docker', and 'Common Metadata for Named Graphs (DCAT)'. Each issue entry includes a title, a 'good first issue' label, and a 'closed' status.



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


Python Apache License 2.0 1 issue needs help Updated yesterday

### BasqueCountryInstitutionsTransparentRelationsGraph Public

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 “escrapear” 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

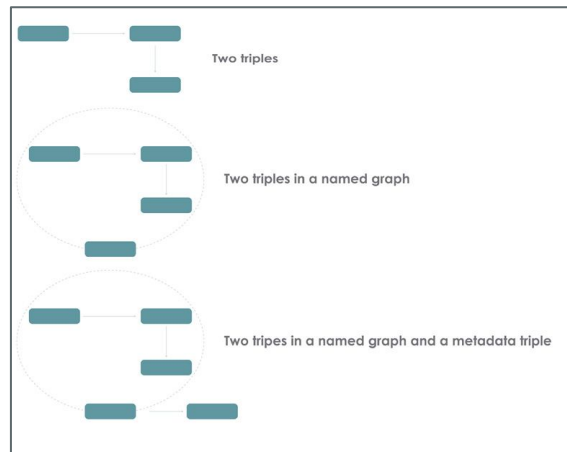
Crear un mapeo sobre un dataset ya existente (ej. MySQL contratos)

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 URIs:

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

## Metadatos:

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

## Formato de fichas de datasets

## Documentación

### *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}`



### *1. URI template/patterns*

ELI's HTTP URIs enable users to access legislation in a persistent way. The URIs 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 URIs, ...



# 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

Posible 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)
- Funciones adicionales: Natural Language Processing (NLP), Named Entity Recognition (NER), reconciliación avanzada de entidades, ChatGPT, ...

# Universidad

The screenshot shows the GitHub interface for the repository 'mikel-egana-aranguren / Transparent-Relations-Ontology'. The repository is public and has a 'develop' branch with 7 branches and 0 tags. The commit history shows several recent commits, including 'Add PPROC as related project' and 'Release 0.1.1'. The README section is visible at the bottom, showing the license as Apache 2.0 and the build tool as Maven 3.6.3.

Commit	Message	Time
mikel-egana-aranguren	Add PPROC as related project	2 weeks ago
mikel-egana-aranguren	Release 0.1.1	last month
mikel-egana-aranguren	Widoco release	last year
mikel-egana-aranguren	Improve annotations	last month
mikel-egana-aranguren	Correct version IRI	last month
mikel-egana-aranguren	Move ROBOT stuff to robot/	last year

**Transparent Relations Ontology**

**KGSWC-2023**  
Fifth Ibero-American Knowledge Graph and Semantic Web Conference joint with  
Fourth Indo-American Knowledge Graph and Semantic Web Conference  
November 13-15, 2023  
University of Zaragoza, Zaragoza, Spain

Proceedings 2022: <https://link.springer.com/book/10.1007/978-3-031-21422-6>

The Knowledge Graph and Semantic Web Conference is an international scientific conference series devoted to knowledge graph and semantic web research. Due to COVID-19 pandemic, this edition of the conference will be hybrid. The prime goals of the conference are:

- (i) to provide a forum for the semantic web community, bringing together researchers and practitioners in industry to share ideas about R&D projects and
- (ii) to increase the adoption of semantic web technologies in the region

Ejercicios

Preguntas

Debate / Trabajo en grupo