

**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

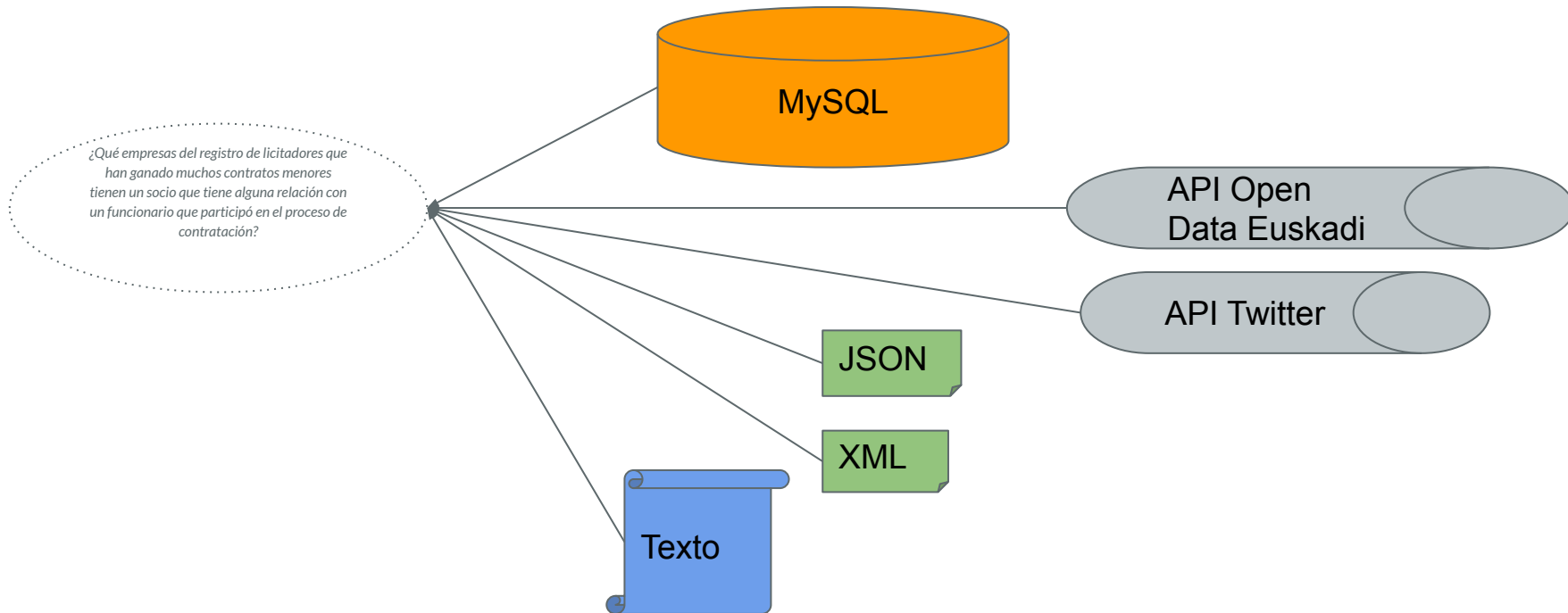
Gran tradición de crítica a las élites y análisis de corrupción: Ahoztar Zelaieta, Hordago, Bilbo Data Lab, ...

Integrar

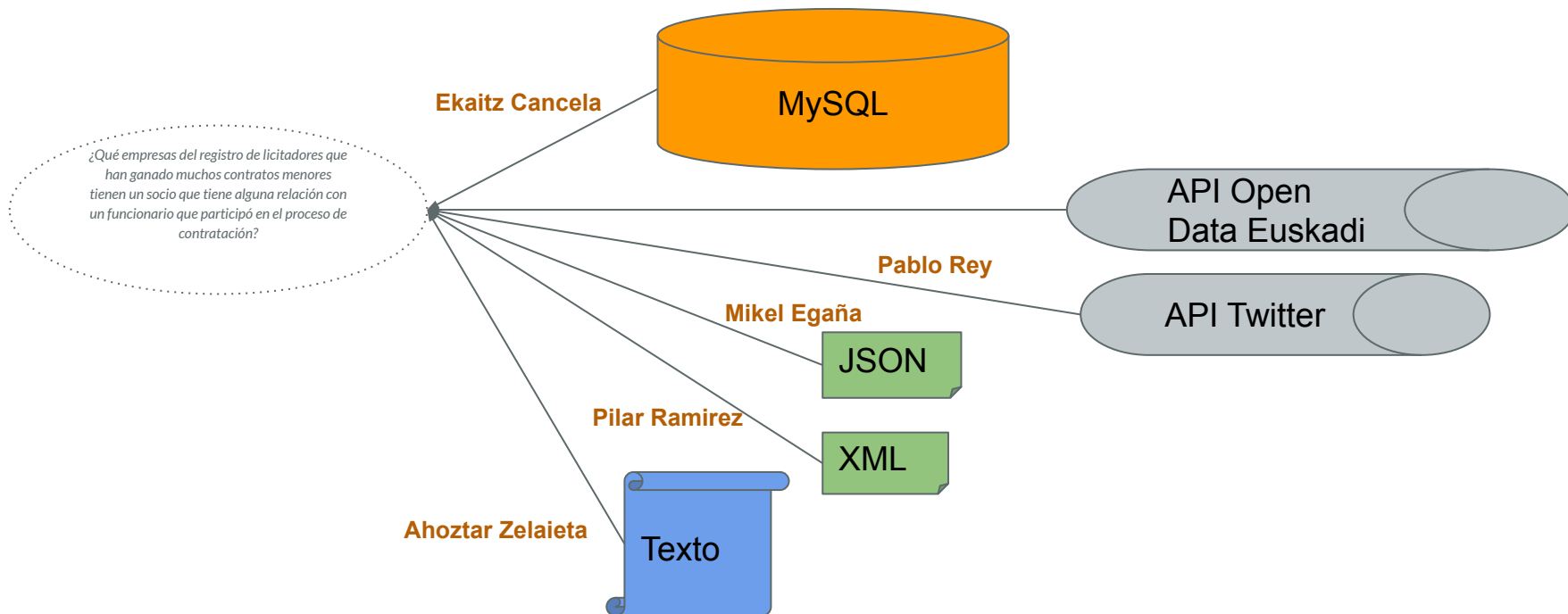
Los datos aislados valen más cuando los integramos con más datos para obtener conclusiones interesantes (Deuda de integración ~ deuda técnica)

¿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?

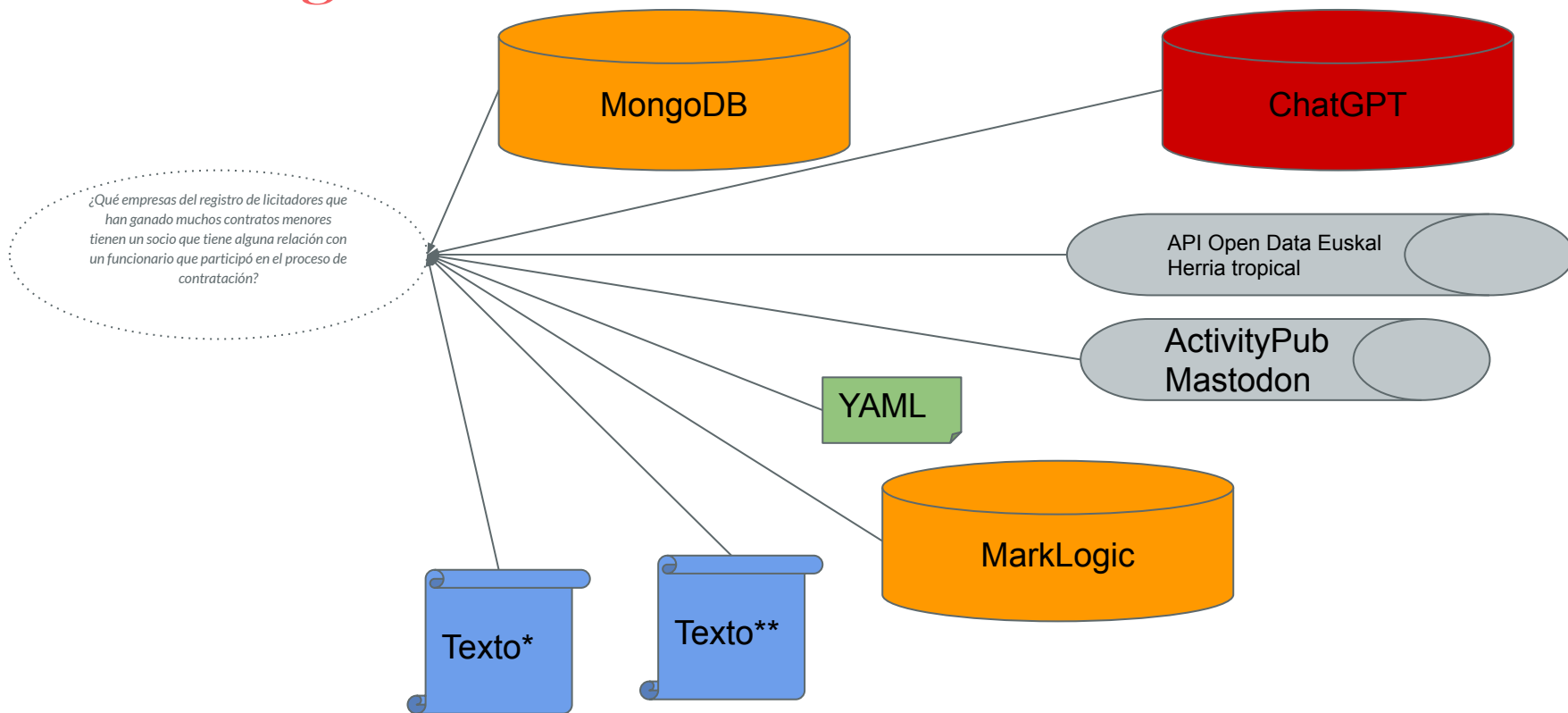
Integrar



De manera colaborativa



Tecnológicamente sostenible



KG

Integrar
De manera colaborativa
Tecnológicamente sostenible

Ekaitz Cancela

MySQL

Pablo Rey

API Open
Data Euskadi

API Twitter

Mikel Egaña

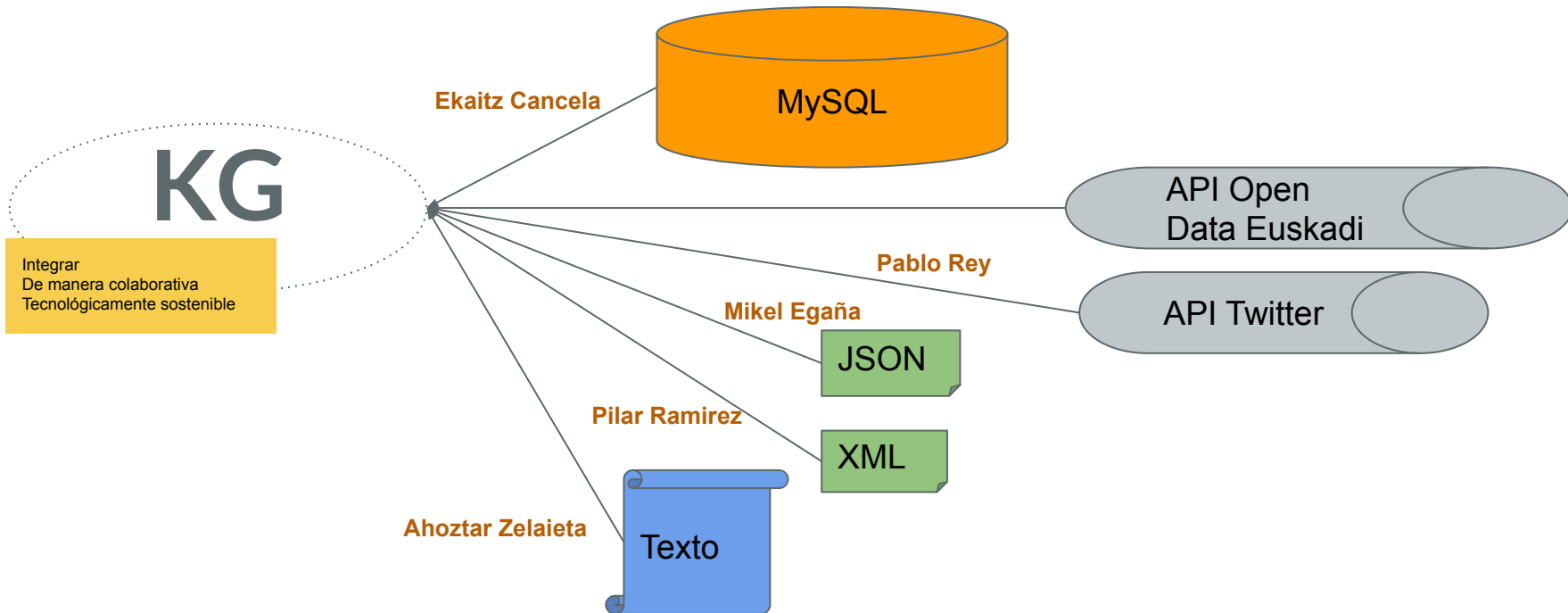
JSON

Pilar Ramirez

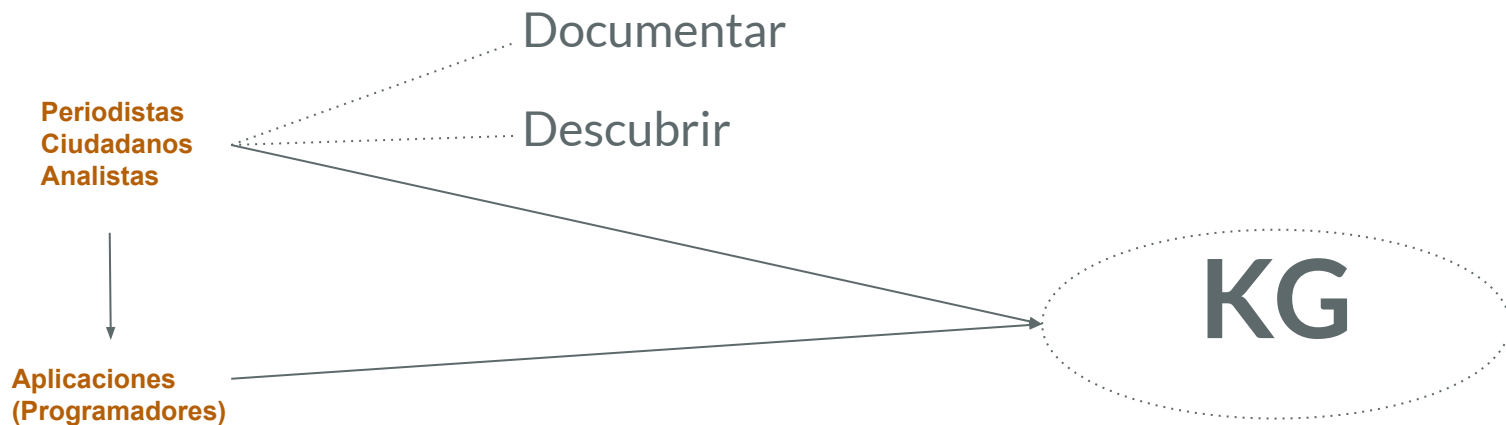
XML

Ahoztar Zelaieta

Texto



Publicar



mikel-egana-aranguren / BasqueCountryInstitutionsTransparentRelationsGraphPublic

<> CodeIssues14Pull requestsDiscussionsActionsProjectsWikiSecurityInsightsSettings

main1 branch0 tagsGo to fileAdd fileCode

mikel-egana-arangurenNo profilesee4e8a5 days ago17 commits

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

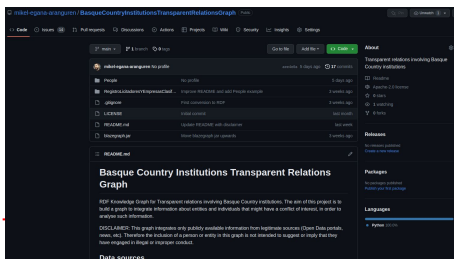
No packages published

Publish your first package

Languages

Python100.0%





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?

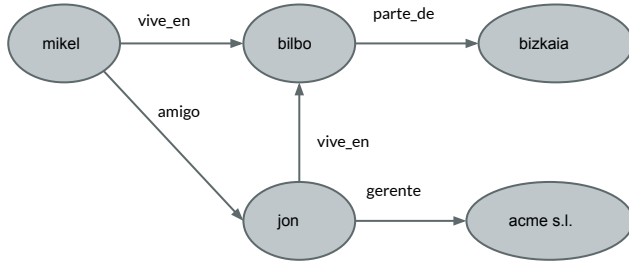
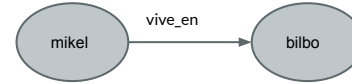
Tecnología

Knowledge Graphs

Graph: Modelo de datos basado en nodos y vértices

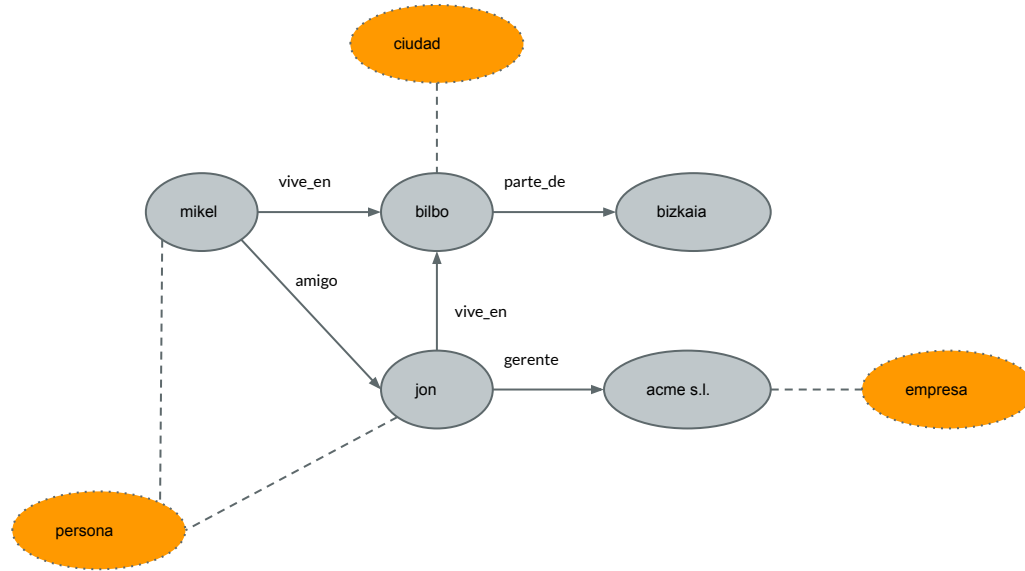
Tripleta formada por dos nodos y un vértice

Muchas tripletas: grafo



Knowledge Graphs

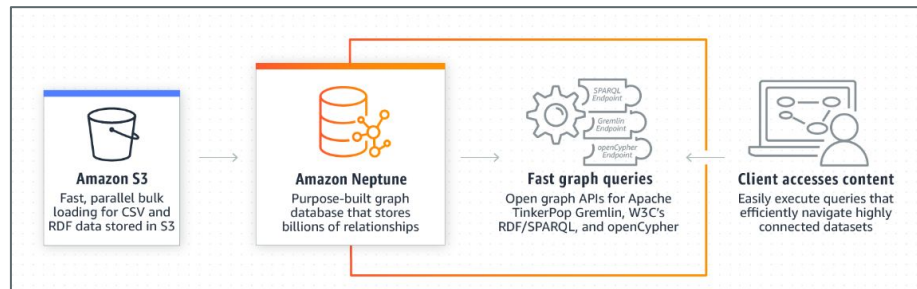
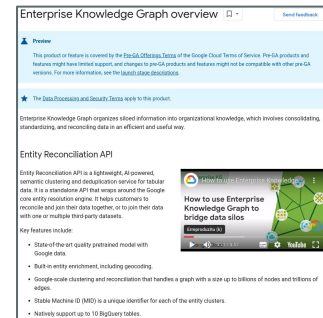
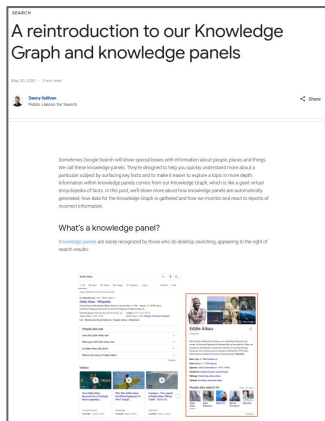
Knowledge: vocabulario común para definir las entidades del grafo



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



Knowledge Graphs



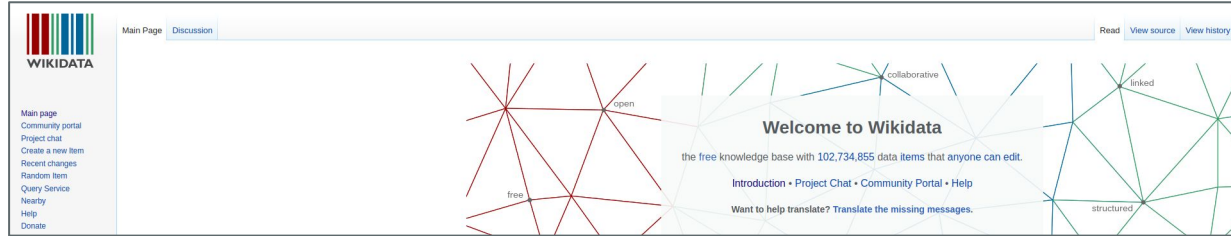
Katariina Kari
Aug 24, 2022 · 5 min read · [Listen](#)

IKEA's Knowledge Graph and Why It Has Three Layers

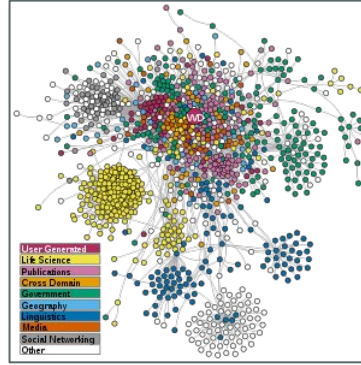
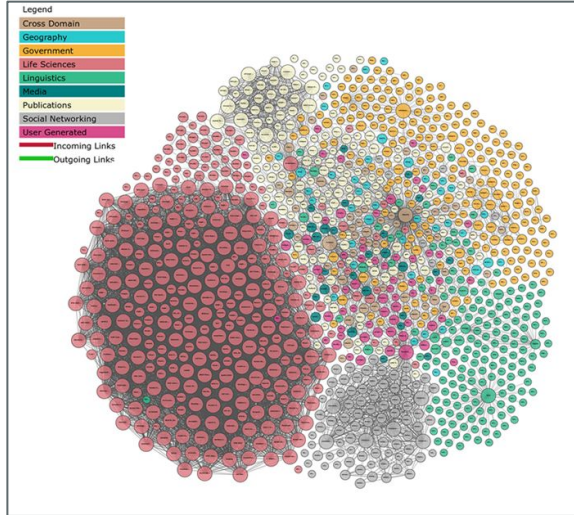
At IKEA we are building a knowledge graph to improve the overall experience of our customers in the physical and digital space. I like to use the following layered pyramid depiction when explaining to stakeholders what a Knowledge Graph (KG) consists of. I saw it for the first time when I watched [this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

```
[mikel@purrutl BilbaoDataLab-Grafos]$ curl -i -H "Accept: application/ld+json" http://www.wikidata.org/entity/Q8692
{
  "@graph": [
    {
      "@id": "data:Q8692",
      "@type": "schema:Dataset",
      "about": "wd:Q8692",
      "license": "http://creativecommons.org/publicdomain/zero/1.0/",
      "softwareVersion": "1.0.0",
      "version": "1869492055",
      "dateModified": "2023-04-06T10:20:23Z",
      "statements": 102,
      "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"
      }
    }
  ]
}
```

Wikidata

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	municipality of Spain
	▼ 0 references
	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:

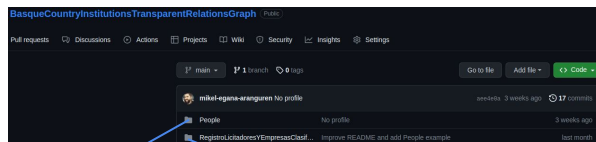
```
<wdtn:P1566 rdf:resource="http://sws.geonames.org/6362368/" />
```

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/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)

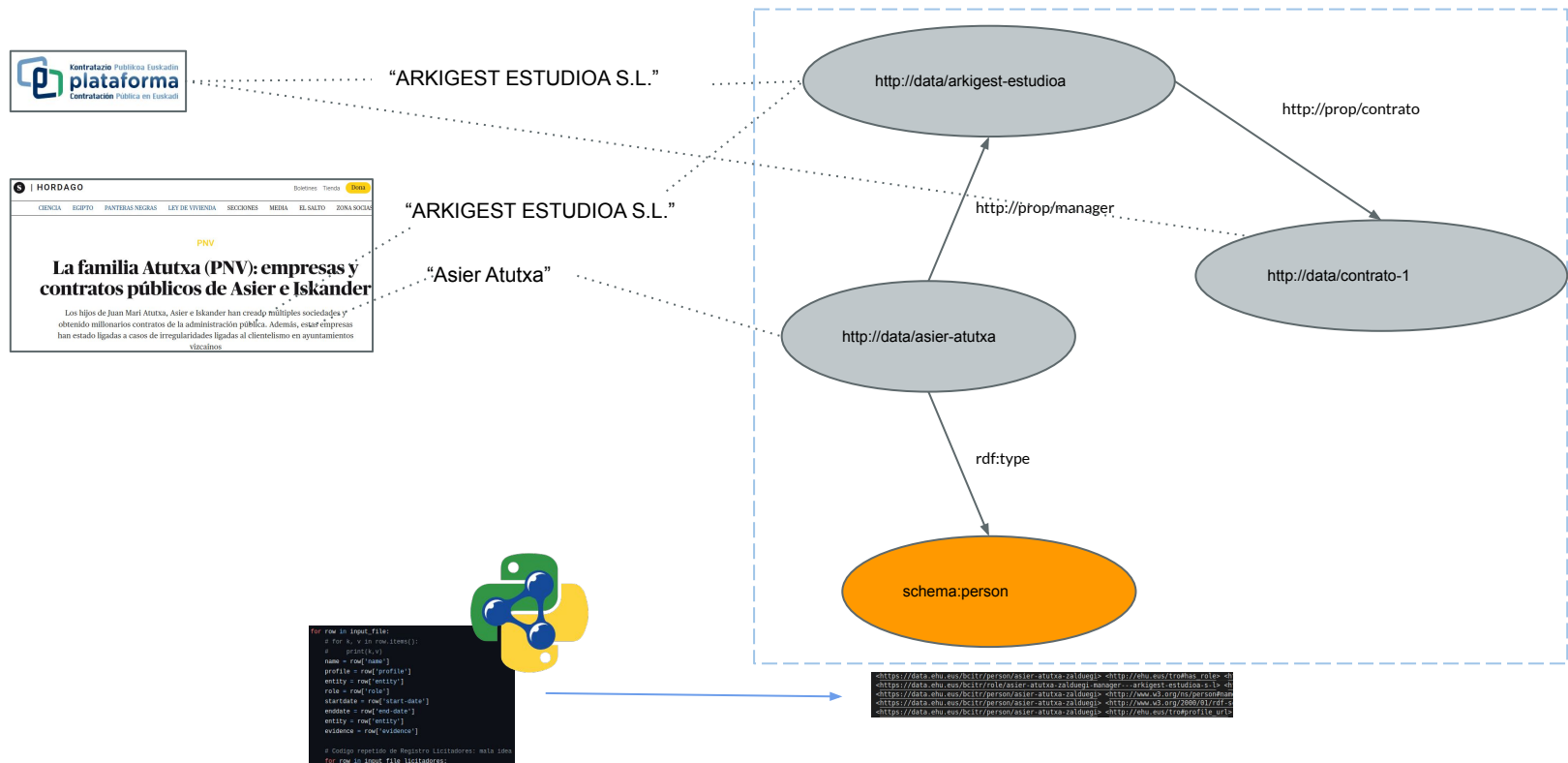
Convert to RDF Graph

- URI pattern:
 - Graph URI: `https://data.ehu.eus/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

main BasqueCountryInstitutionsTransparentRelationsGraph / RegistroLicitadoresYEmpresasClasificadas /

mikel-egana-aranguren Improve README and add People example

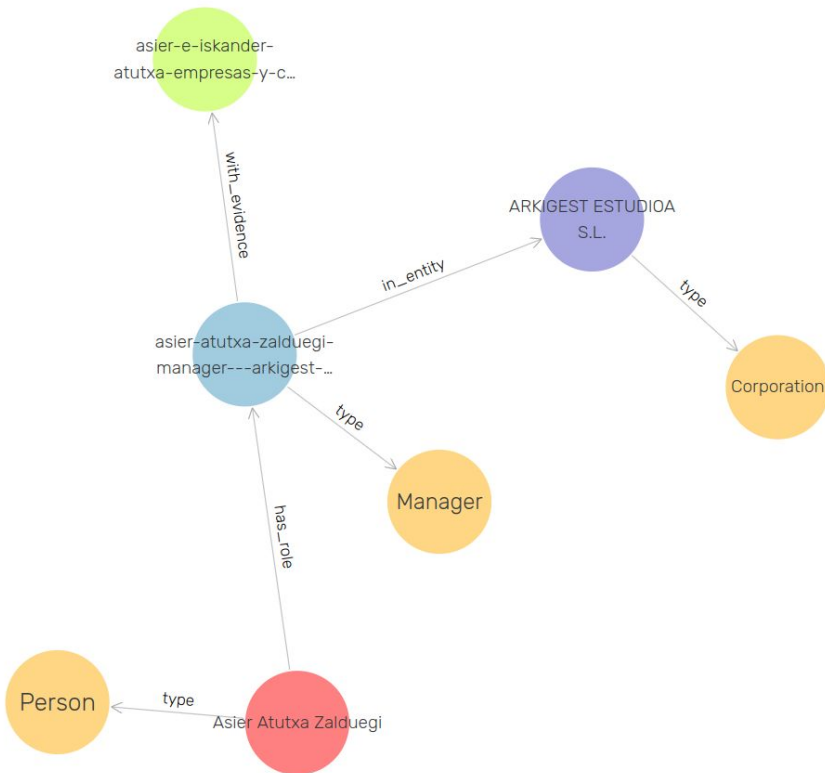
..

CSV2RDF.py	improve licitadores
Empresas-2023-03-05-metadata.ttl	improve licitadores
Empresas-2023-03-05.csv	First conversion to RDF
Empresas-2023-03-05.nq	type

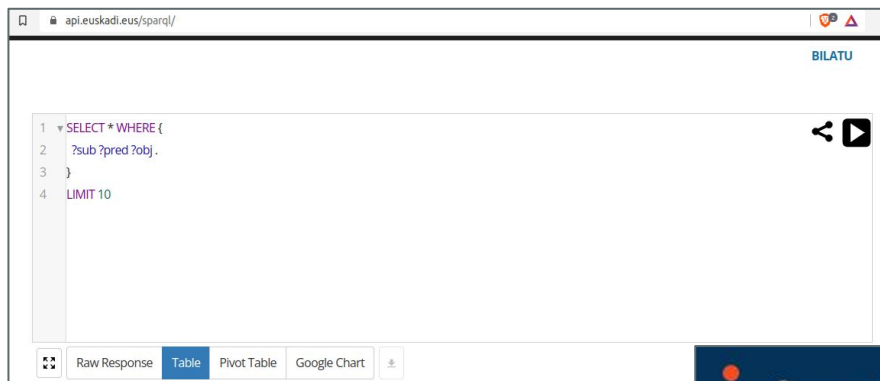
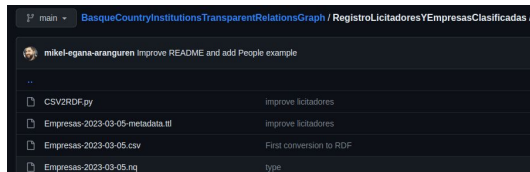
```
<https://data.ehu.es/bcitr/person/asier-atutxa-zalduegi> <http://ehu.es/irohbas/role/>  
<https://data.ehu.es/bcitr/role/asier-atutxa-zalduegi-manager> <arkquest-estudios> <1> <  
<https://data.ehu.es/bcitr/person/asier-atutxa-zalduegi> <http://www.w3.org/ns/person>  
<https://data.ehu.es/bcitr/person/asier-atutxa-zalduegi> <http://www.w3.org/2000/01/rdf-schema#>  
<https://data.ehu.es/bcitr/person/asier-atutxa-zalduegi> <http://ehu.es/irohbas/role/>
```



Triple Store (GraphDB local)

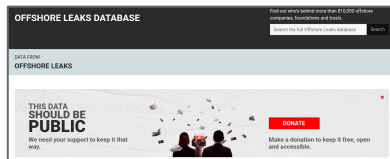
[illegible]

4) Publicación

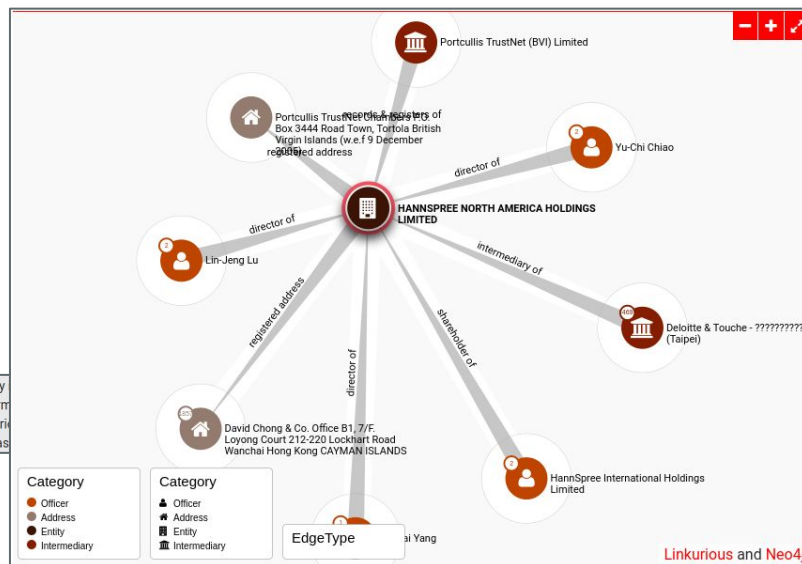


Otros proyectos

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 ICIJ has received in connection with various database. 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:

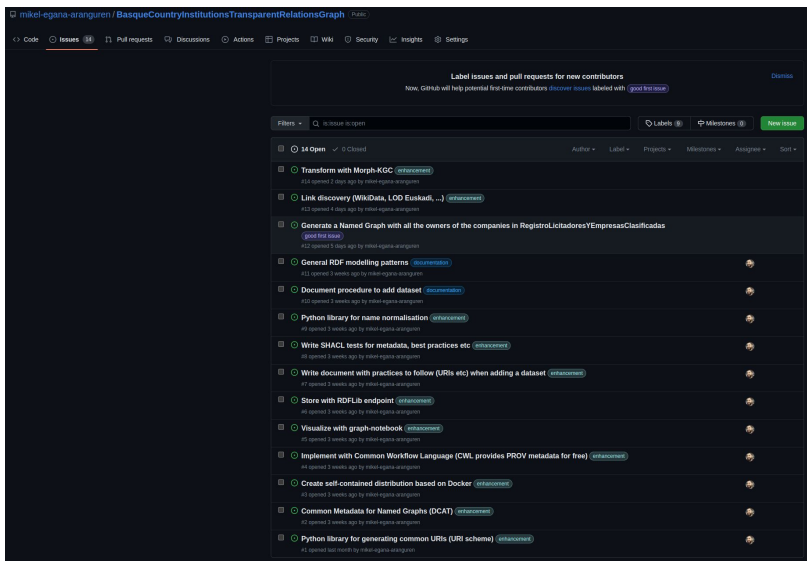
- Nos permiten centrarnos en los problemas de integración
- Nos permiten reusar el trabajo de integración

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



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](#)

[Type](#) [Language](#) [Sort](#) [New](#)

1 result for repositories matching **Basque** sorted by **last updated** [Clear filter](#)

BasqueCountryInstitutionsTransparentRelationsGraph [Public](#) [Star](#)

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](#)

[bd1](#) [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

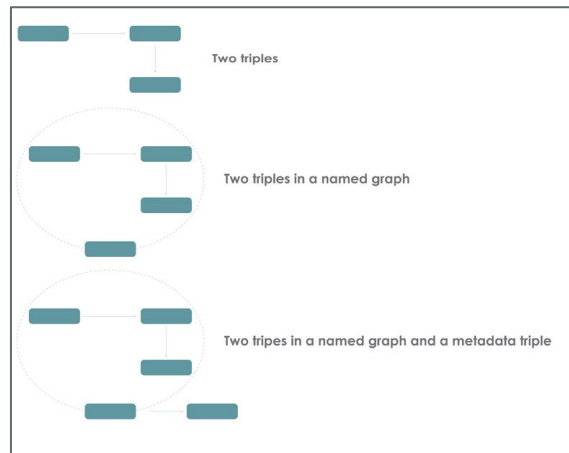
Crear un mapeo sobre un dataset ya existente (ej. MySQL contratos)

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

mikel-egana-aranguren / Transparent-Relations-Ontology

Code Issues 5 Pull requests Discussions Actions Projects Wiki Security Insights Settings

develop 7 branches 0 tags

Go to file Add file + Code +

mikel-egana-aranguren Add PPROC as related project 1488938 2 weeks ago 28 commits

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 Maven C1

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

